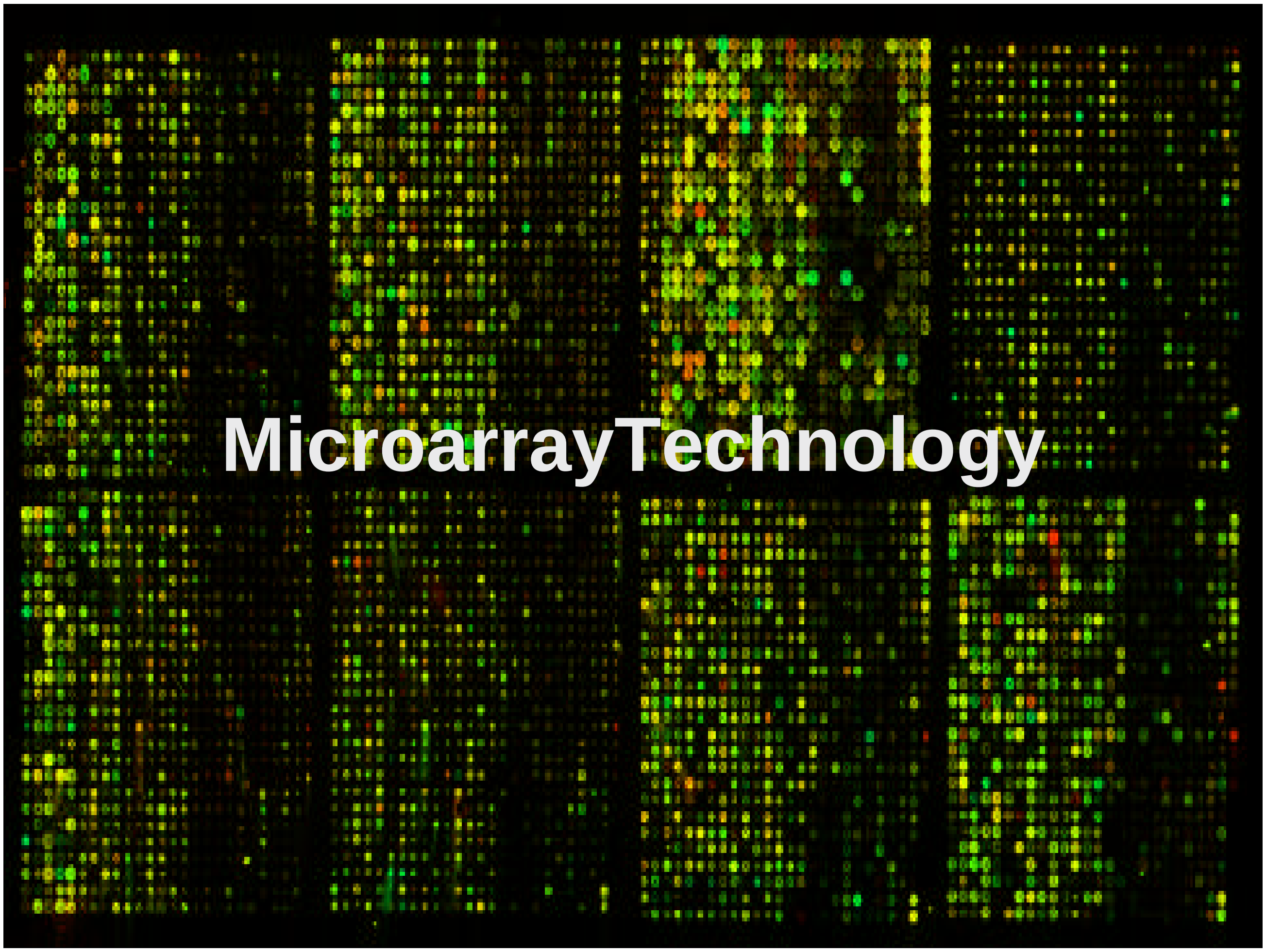
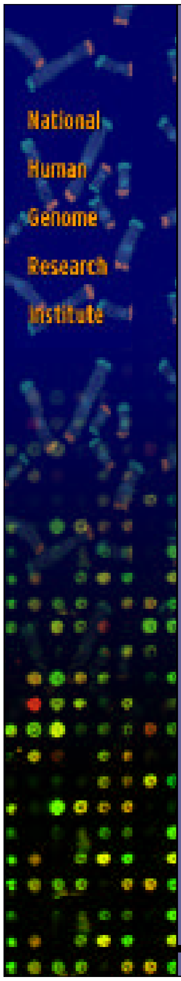


MicroarrayTechnology



AFTER THE SEQUENCE:

WHOLE GENOME APPROACHES TO

BIOLOGICAL QUESTIONS

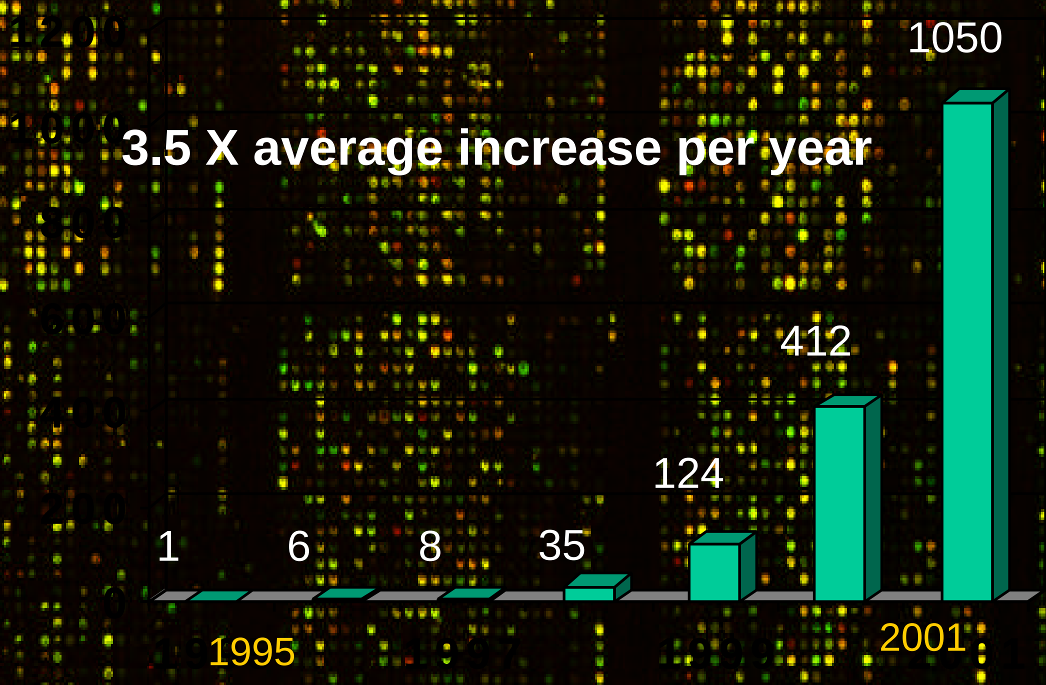
GENE EXPRESSION

GENE VARIATION

GENE FUNCTION

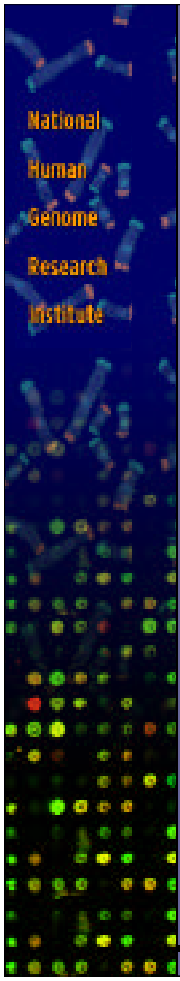
Cancer Genetics Branch

PUBMED literature on DNA microarrays



3.5 X average increase per year

No. of publications per year



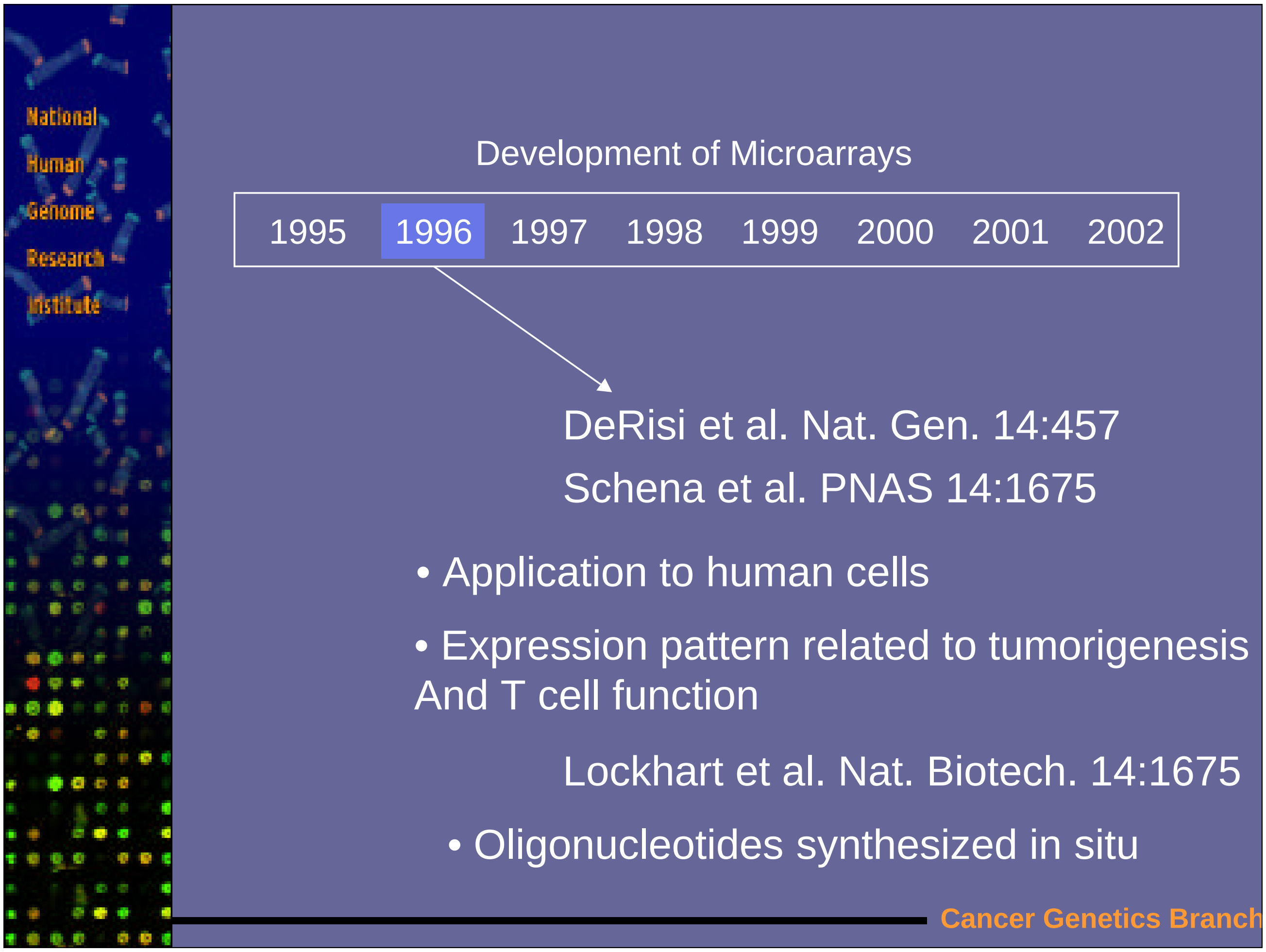
Development of Microarrays

1995
1996
1997
1998
1999
2000
2001
2002

Schena et al. Science 270:467

- Robotic high density printing of cDNAs
- Fluorescence detection

Cancer Genetics Branch



National
Human
Genome
Research
Institute

Development of Microarrays

1995 1996 1997 1998 1999 2000 2001 2002

DeRisi et al. Nat. Gen. 14:457

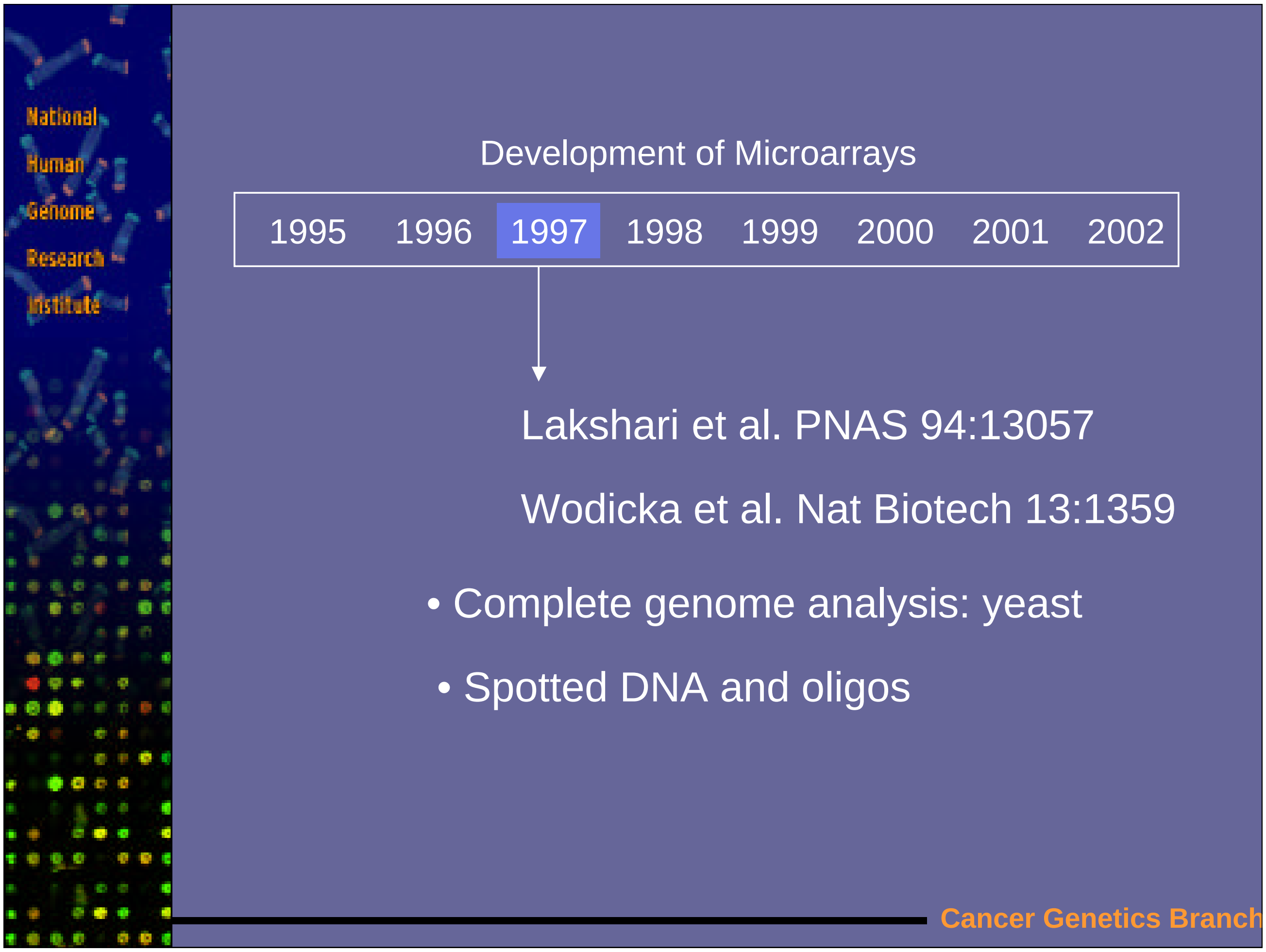
Schena et al. PNAS 14:1675

- Application to human cells
- Expression pattern related to tumorigenesis And T cell function

Lockhart et al. Nat. Biotech. 14:1675

- Oligonucleotides synthesized in situ

Cancer Genetics Branch



National
Human
Genome
Research
Institute

Development of Microarrays

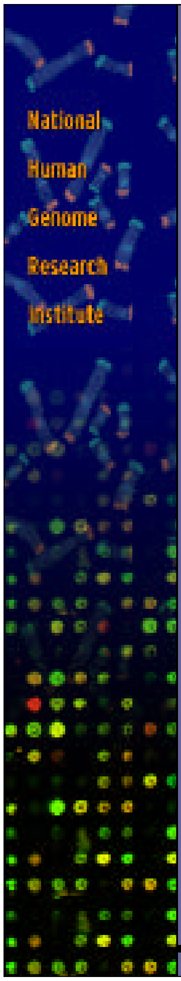
1995 1996 1997 1998 1999 2000 2001 2002

Lakshari et al. PNAS 94:13057

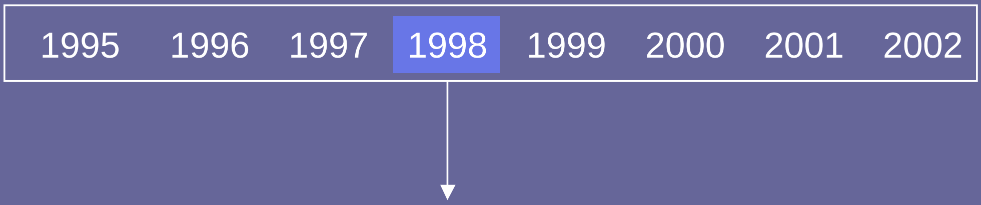
Wodicka et al. Nat Biotech 13:1359

- Complete genome analysis: yeast
- Spotted DNA and oligos

Cancer Genetics Branch



Development of Microarrays for Cancer Research



Khan et al. Cancer Res 58:5009

- Cancers of the same type cluster.

Eisen et al. PNAS 95:14863

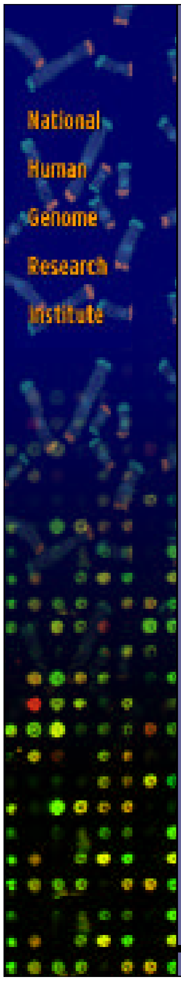
- Two dimensional clustering.

Kononen et al. Nat Med 4:844

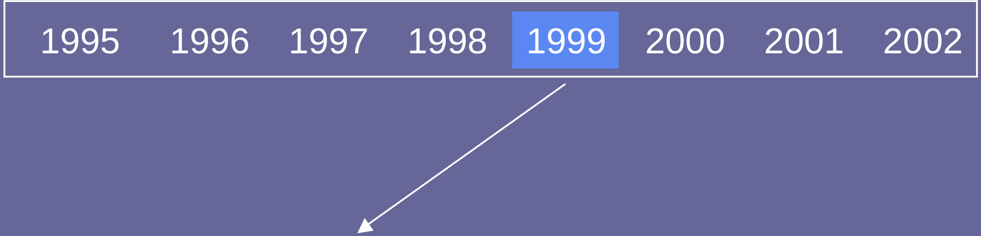
- Tissue microarrays.

Pinkel et al. Nat Gen 20:207

- CGH BAC arrays.



Development of Microarrays



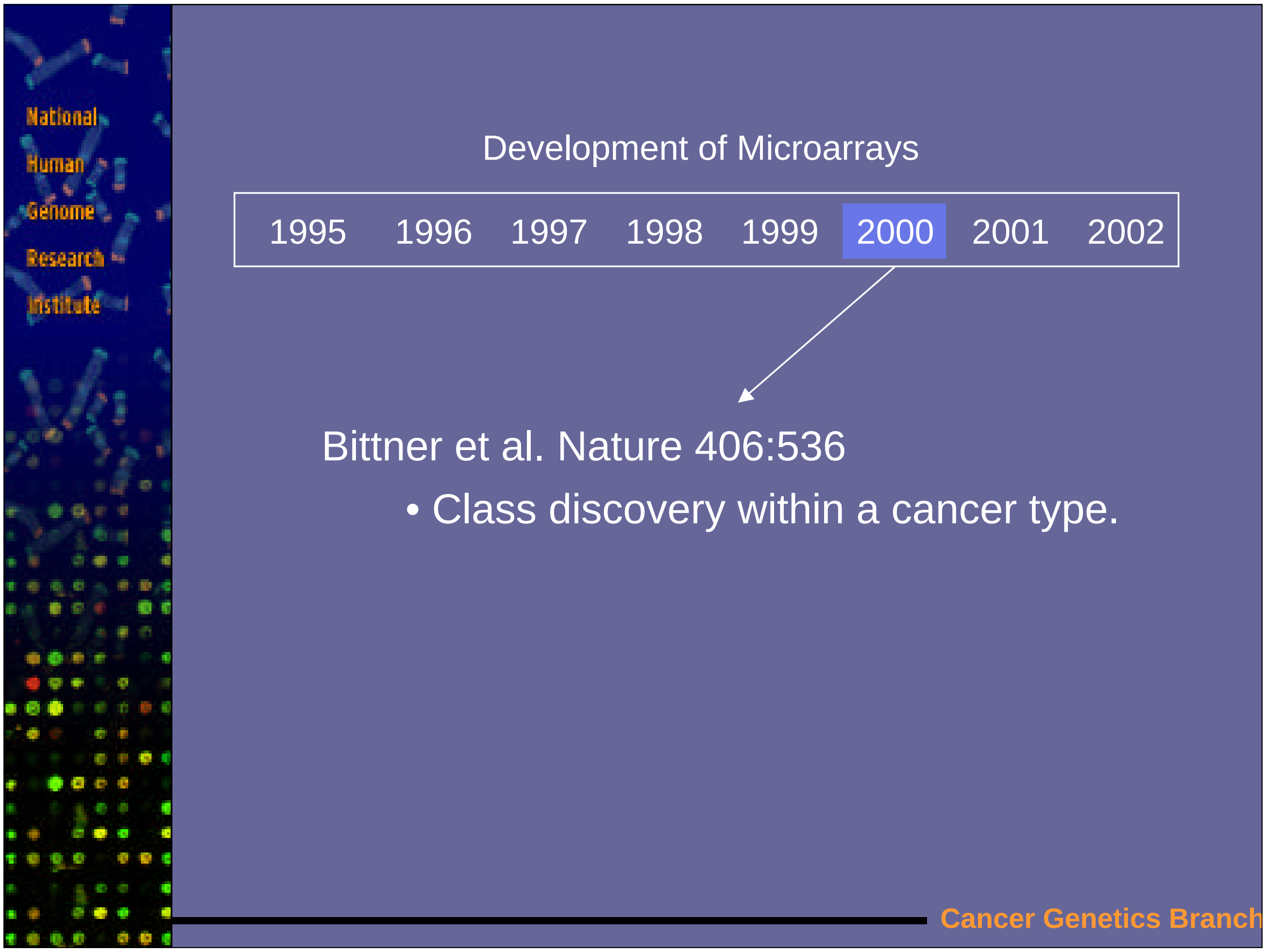
Khan et al. PNAS 96:13464

- Expression program elicited by oncogene.

Golub et al. PNAS 286:531

- Formal diagnostic classifier.

Several sample clustering papers.



National
Human
Genome
Research
Institute

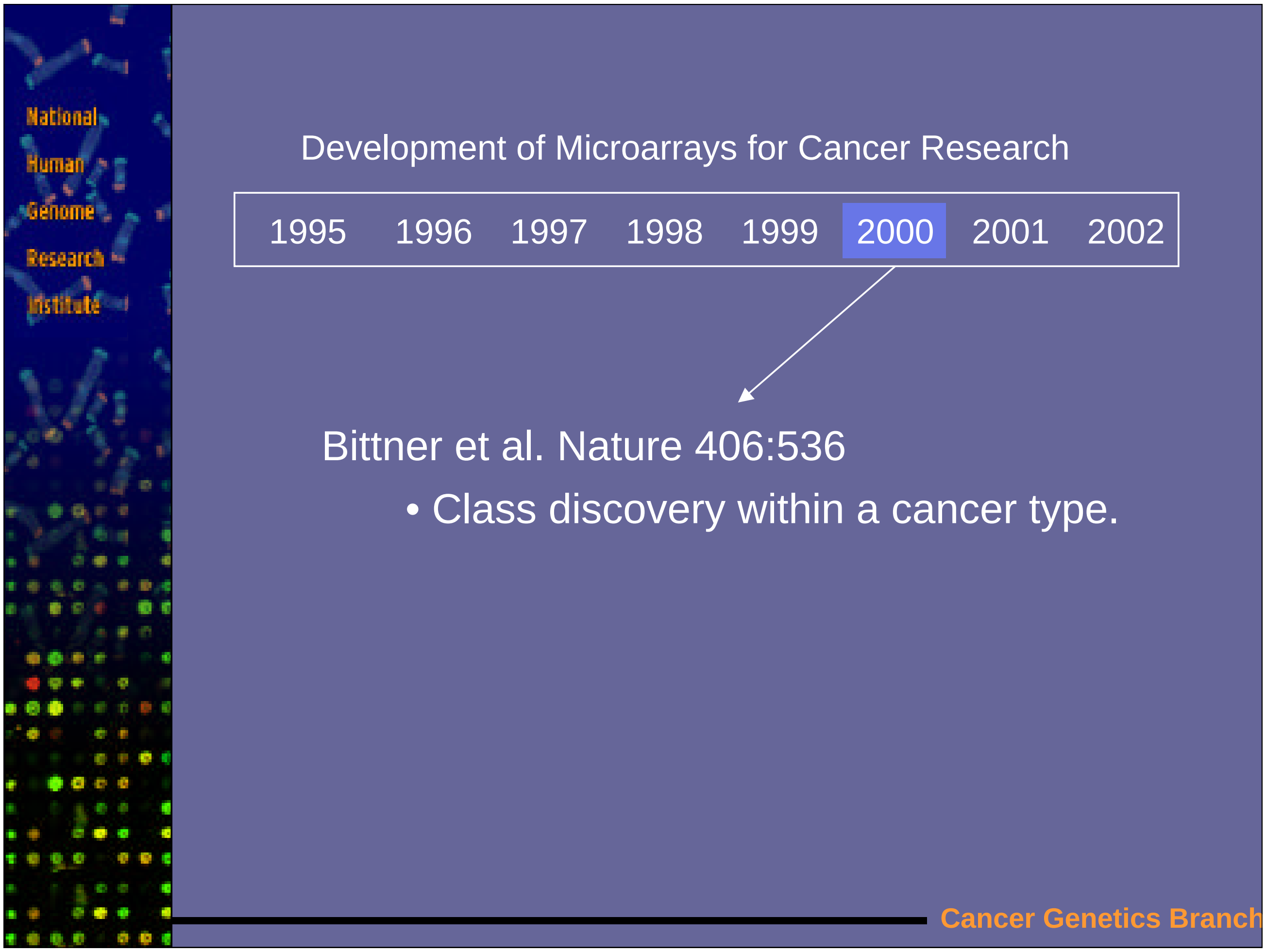
Development of Microarrays

1995 1996 1997 1998 1999 2000 2001 2002

Bittner et al. Nature 406:536

- Class discovery within a cancer type.

Cancer Genetics Branch



National
Human
Genome
Research
Institute

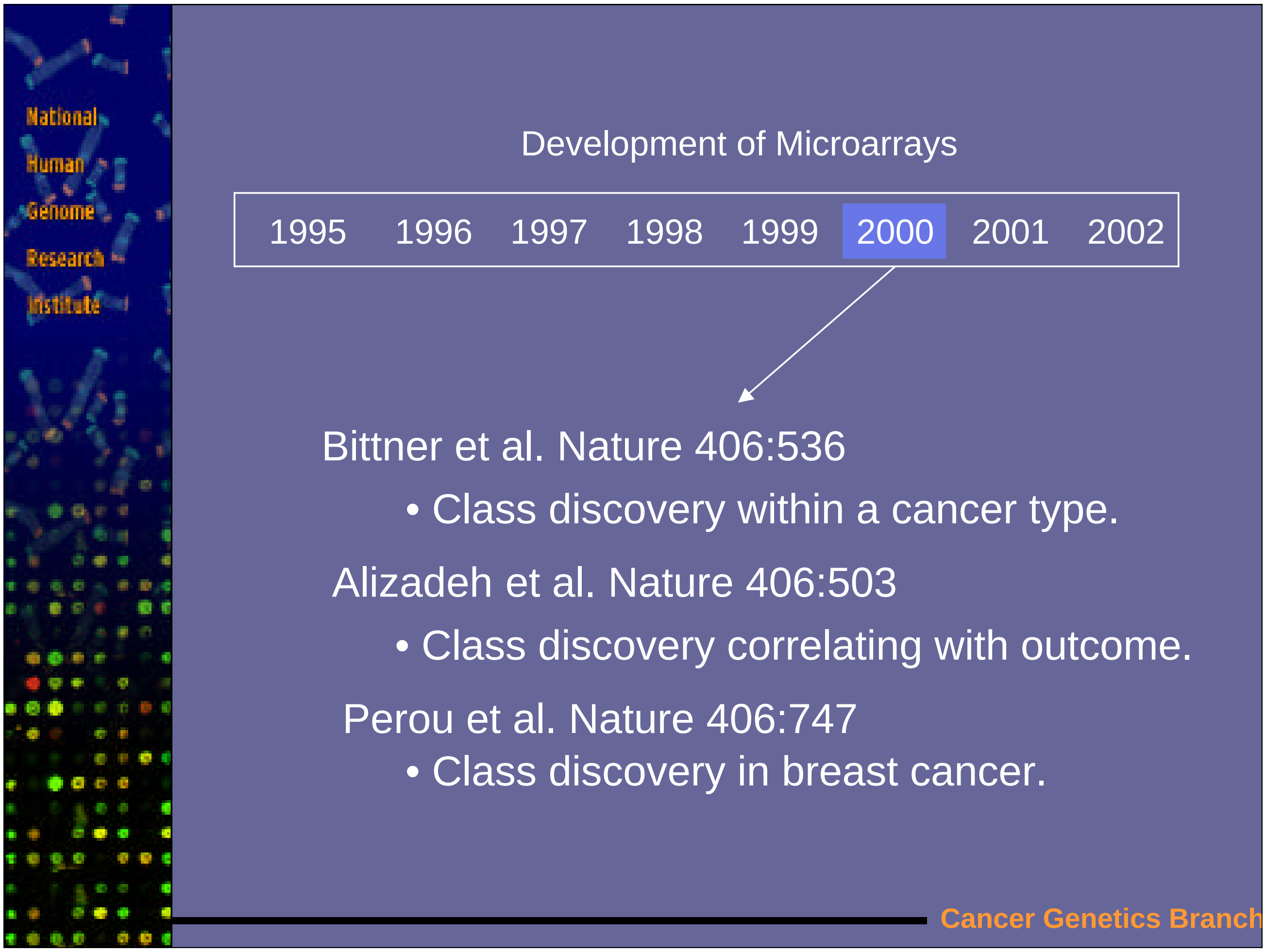
Development of Microarrays for Cancer Research

1995 1996 1997 1998 1999 2000 2001 2002

Bittner et al. Nature 406:536

- Class discovery within a cancer type.

Cancer Genetics Branch



National Human Genome Research Institute

Development of Microarrays

1995 1996 1997 1998 1999 **2000** 2001 2002

Bittner et al. Nature 406:536

- Class discovery within a cancer type.

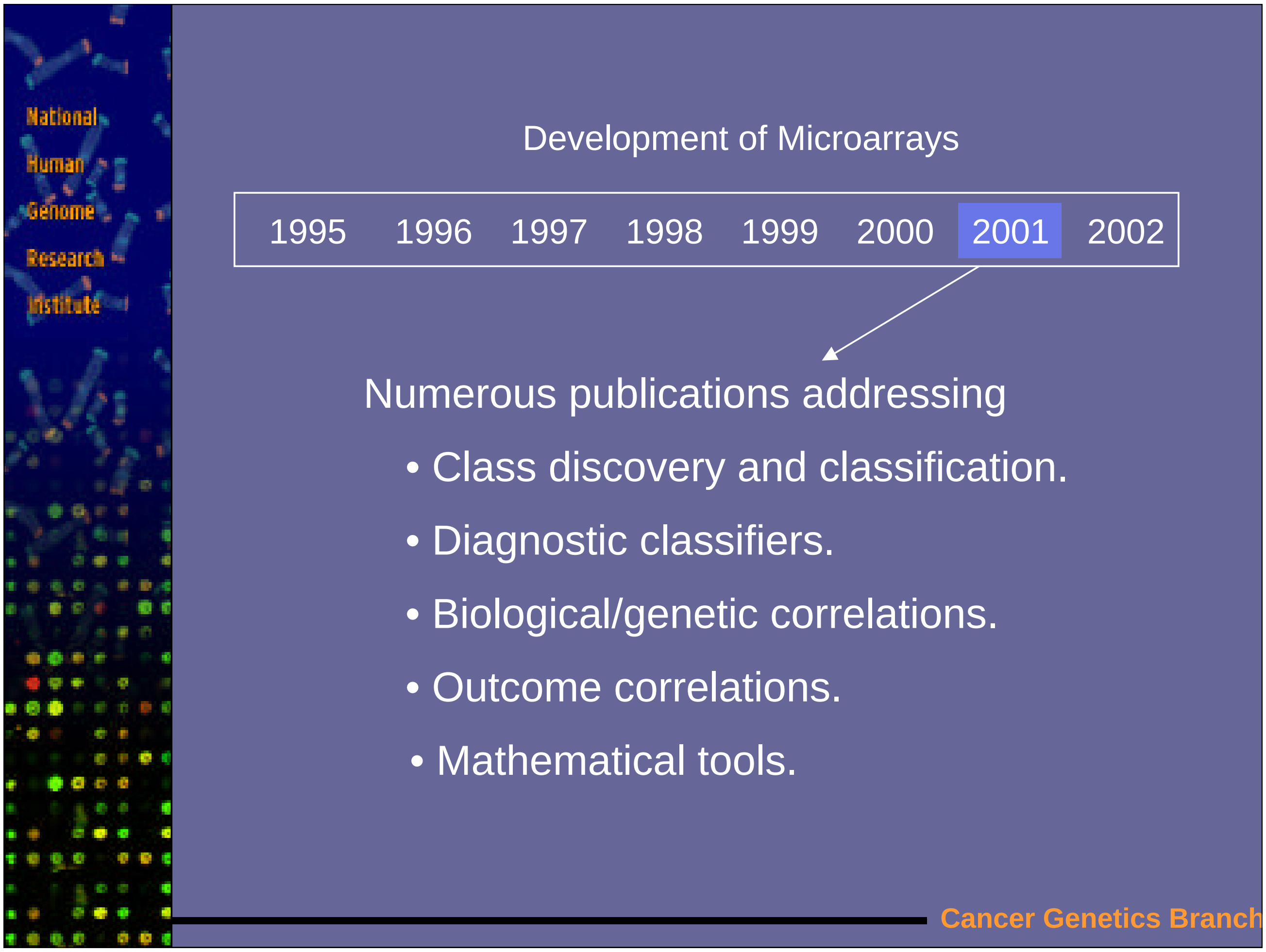
Alizadeh et al. Nature 406:503

- Class discovery correlating with outcome.

Perou et al. Nature 406:747

- Class discovery in breast cancer.

Cancer Genetics Branch



National Human Genome Research Institute

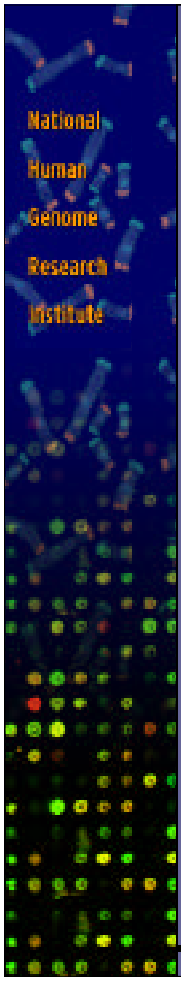
Development of Microarrays

1995 1996 1997 1998 1999 2000 **2001** 2002

Numerous publications addressing

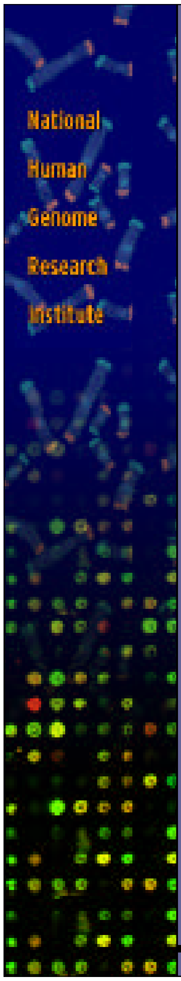
- Class discovery and classification.
- Diagnostic classifiers.
- Biological/genetic correlations.
- Outcome correlations.
- Mathematical tools.

Cancer Genetics Branch



MICROARRAY TERMINOLOG

- Feature--an array element
- Probe--a feature corresponding to a defined sequence
- Target--a pool of nucleic acids of unknown sequence



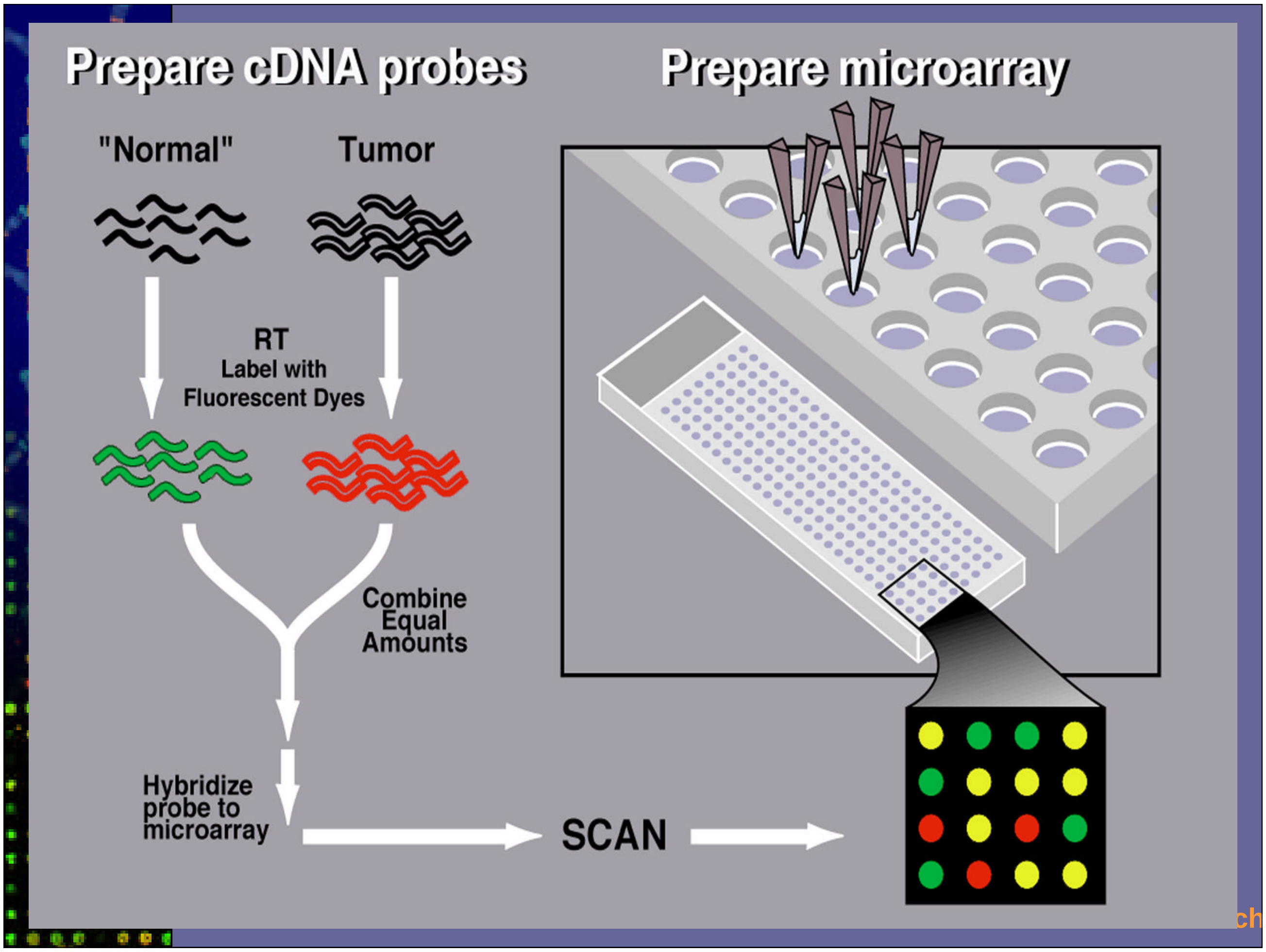
Kinds of array elements

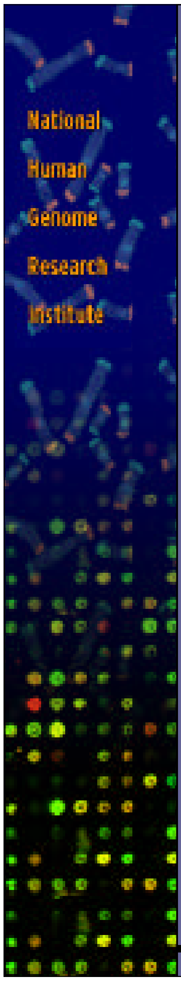
- Synthetic Oligonucleotides
- PCR products from
  - Cloned DNAs
  - Genomic DNA
- Cloned DNA

National
Human
Genome
Research
Institute

Microarray Manufacture

- Printing

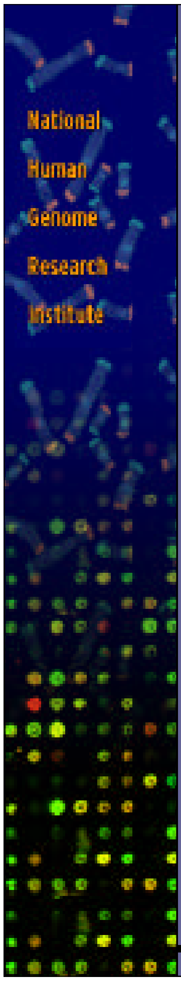




Microarray Manufacture

- Printing

- Synthesis *in situ*



MICROARRAY READOUT

•Determine quantity of target bound to each probe in a complex hybridization

•Must have high sensitivity, low background

•High spatial resolution essential

•Dual channel capability preferred

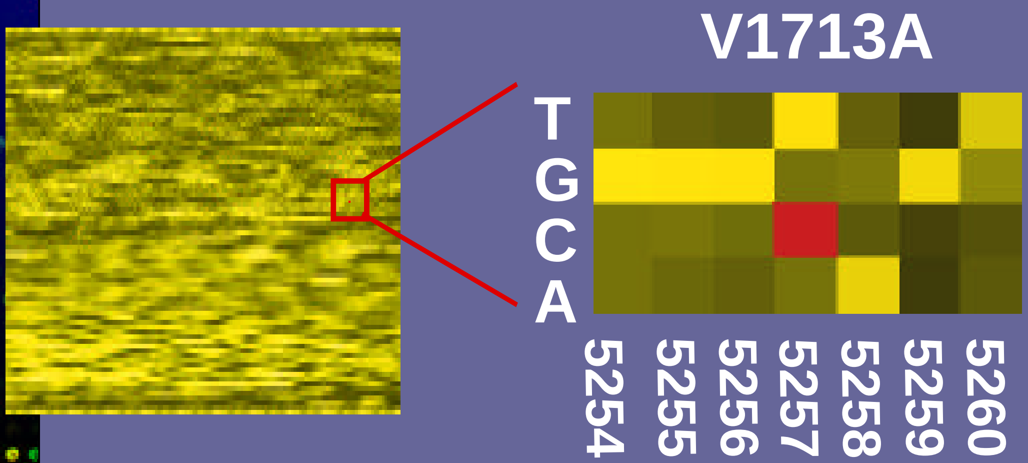
DNA Microarray Applications

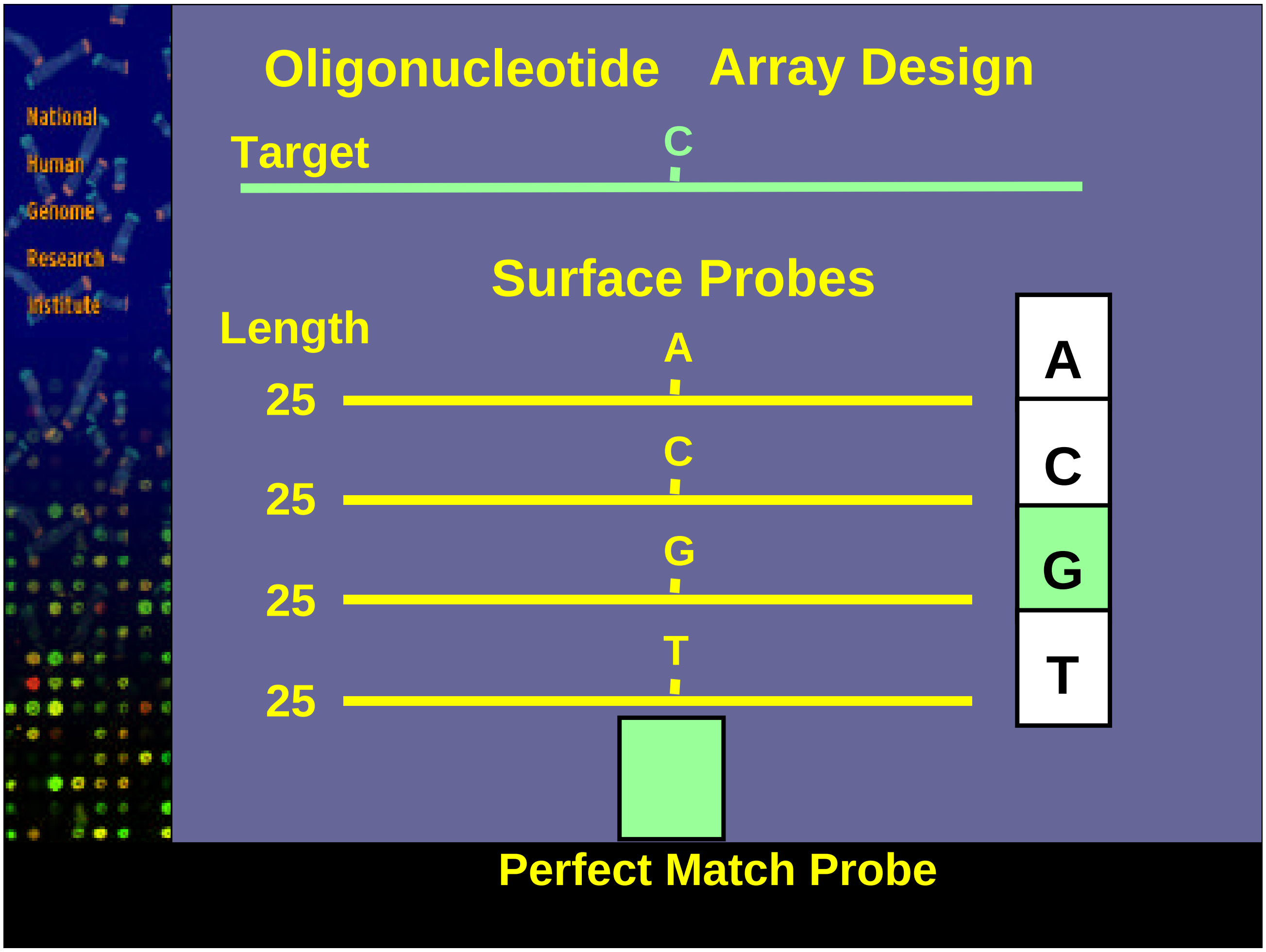
- Resequencing

  Mutations

  Polymorphisms

BRCA1 Coding Region Array

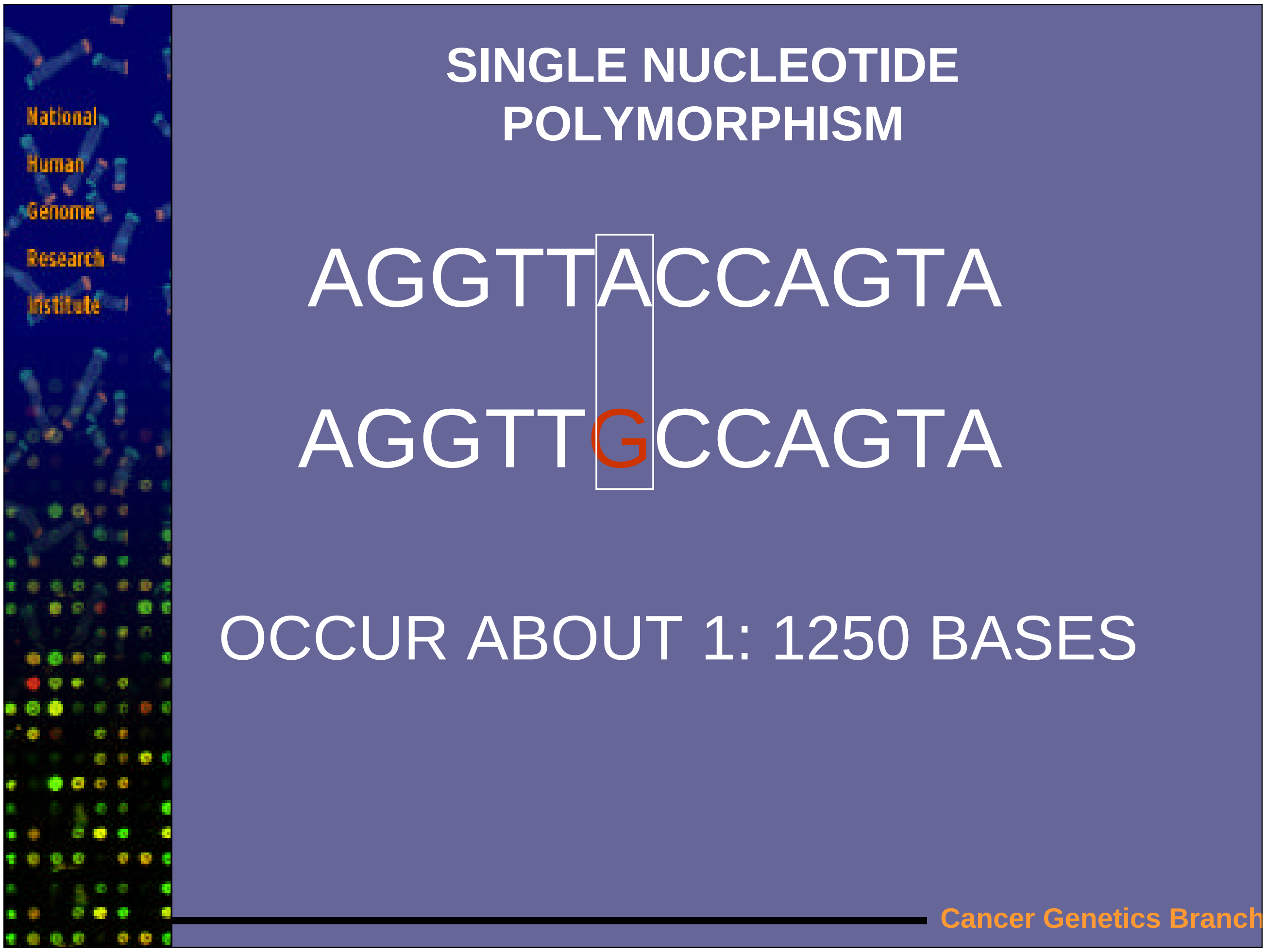




National
Human
Genome
Research
Institute

Oligonucleotide Array Design

Perfect Match Probe



National
Human
Genome
Research
Institute

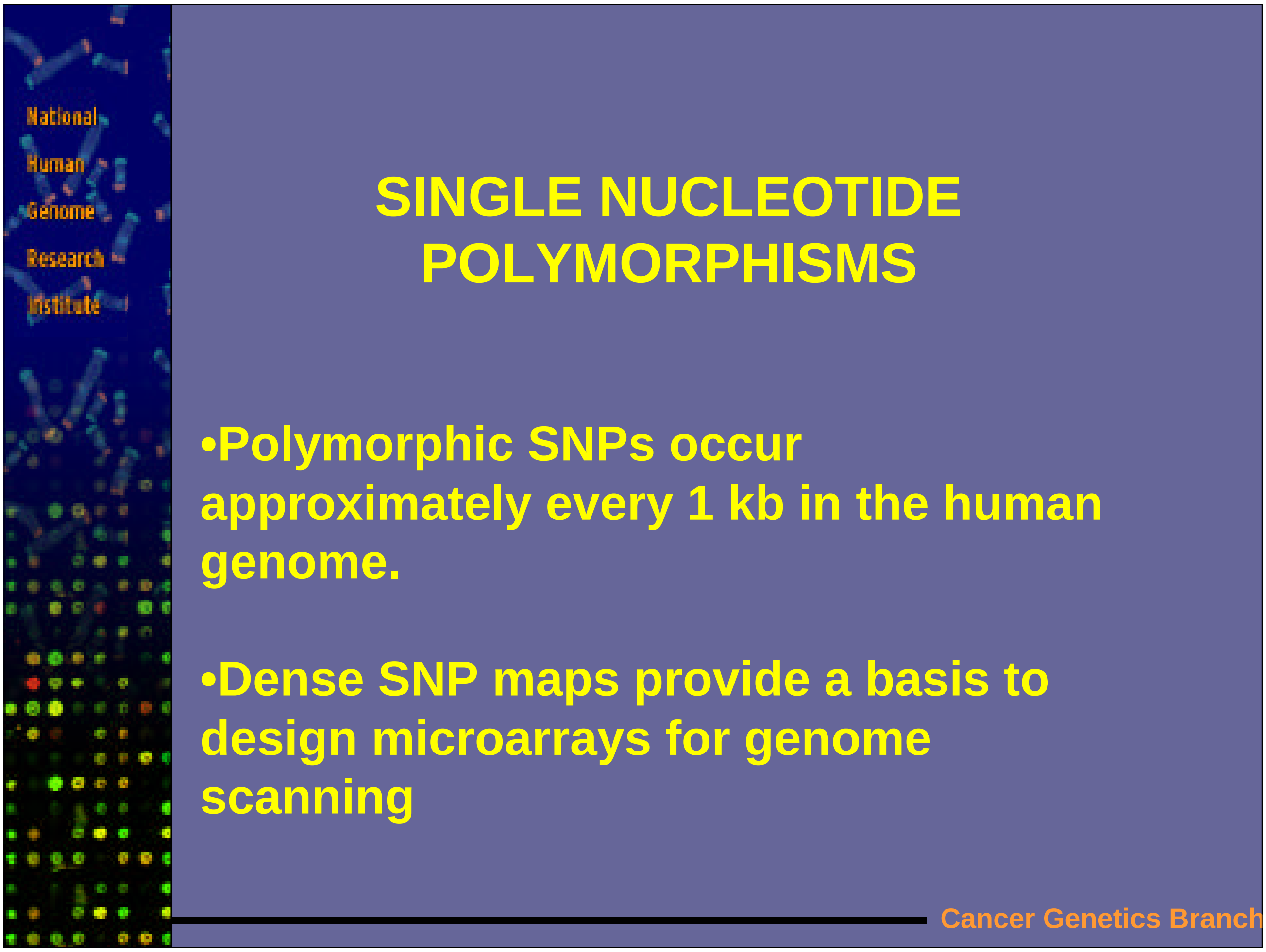
SINGLE NUCLEOTIDE POLYMORPHISM

AGGTTACCAGTA

AGGTTGCCAGTA

OCCUR ABOUT 1: 1250 BASES

Cancer Genetics Branch

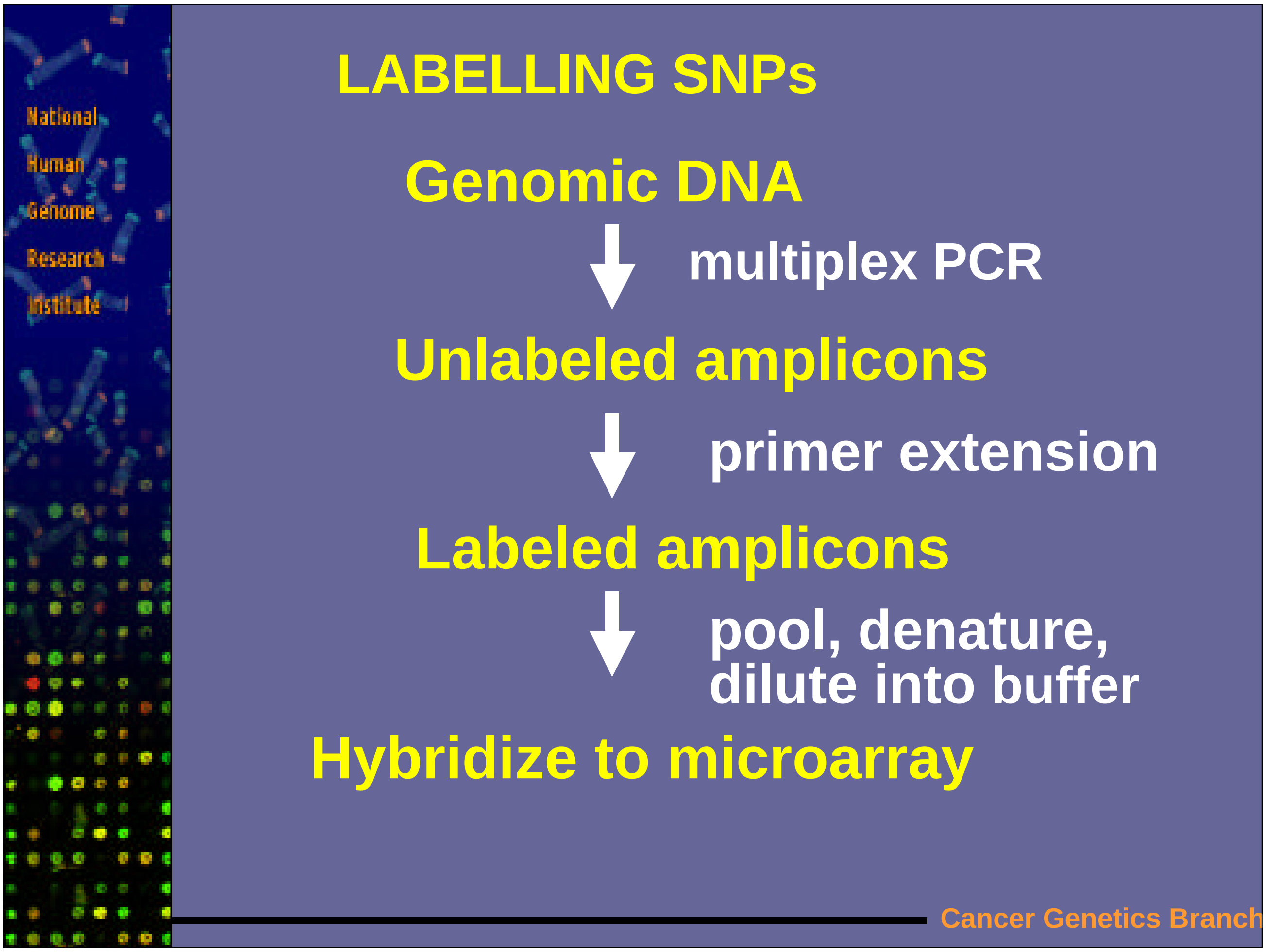


National
Human
Genome
Research
Institute

SINGLE NUCLEOTIDE POLYMORPHISMS

- Polymorphic SNPs occur approximately every 1 kb in the human genome.
- Dense SNP maps provide a basis to design microarrays for genome scanning

Cancer Genetics Branch



National
Human
Genome
Research
Institute

LABELLING SNPs

Genomic DNA

↓ multiplex PCR

Unlabeled amplicons

↓ primer extension

Labeled amplicons

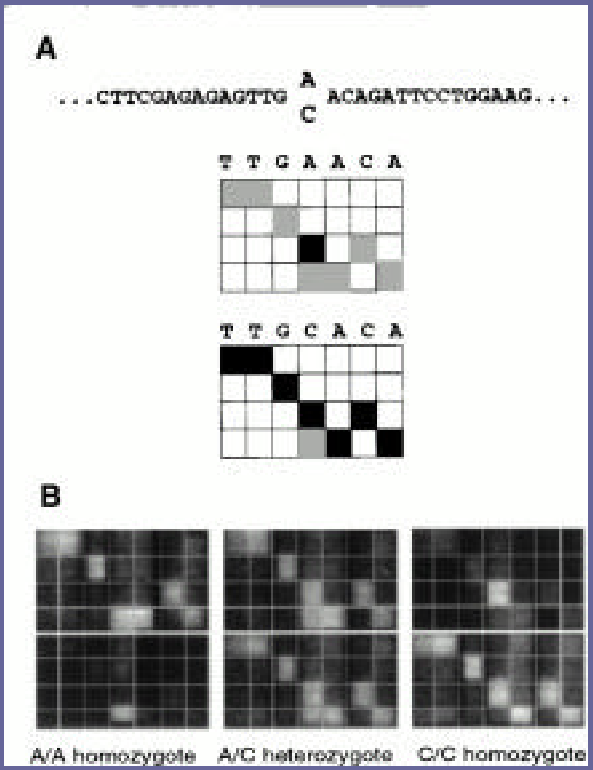
↓ pool, denature,
dilute into buffer

Hybridize to microarray

Cancer Genetics Branch

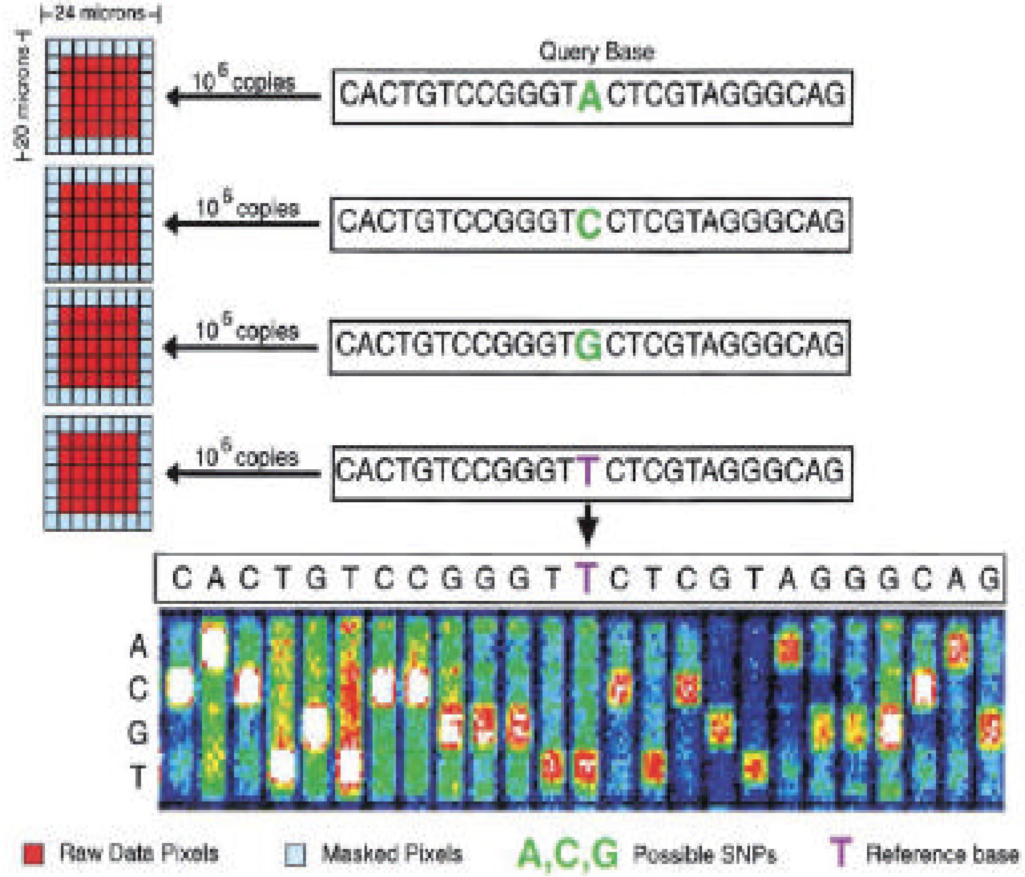
National
Human
Genome
Research
Institute

SNP CHIP*



*Wang et al.
Science 280:1077
1998

National
Human
Genome
Research
Institute



Cutler DJ et al. Genome Res. 2001 11:1913-25.

National
Human
Genome
Research
Institute

ACCURACY OF SNP CHIP

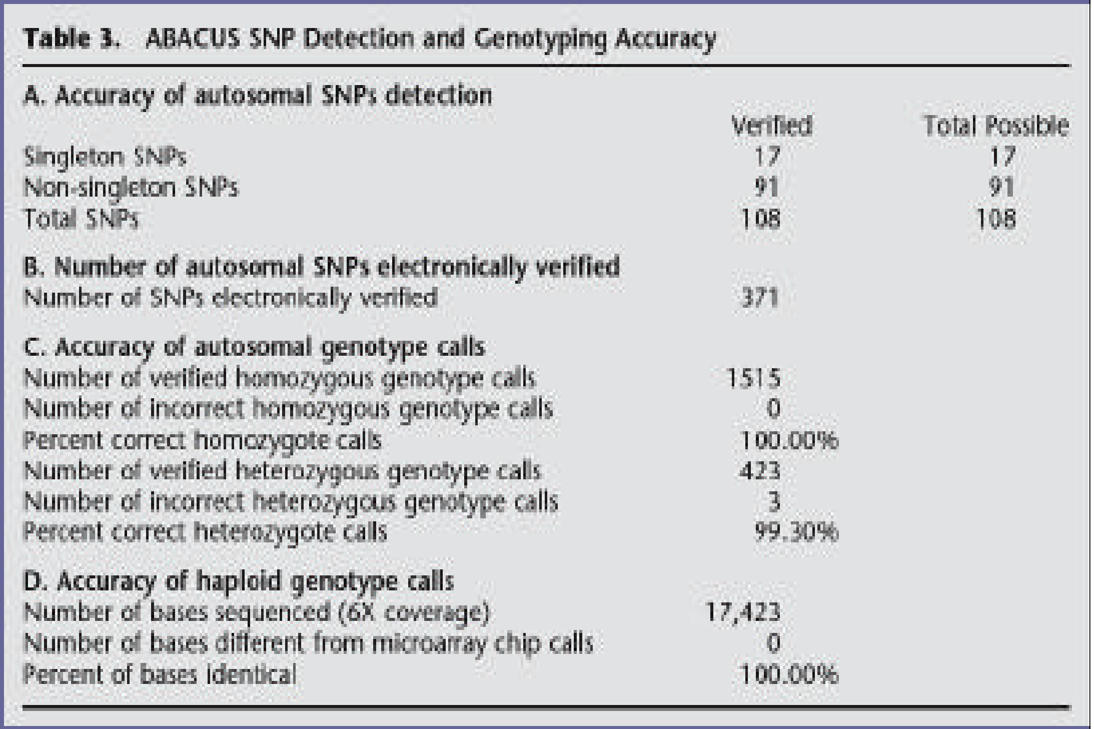


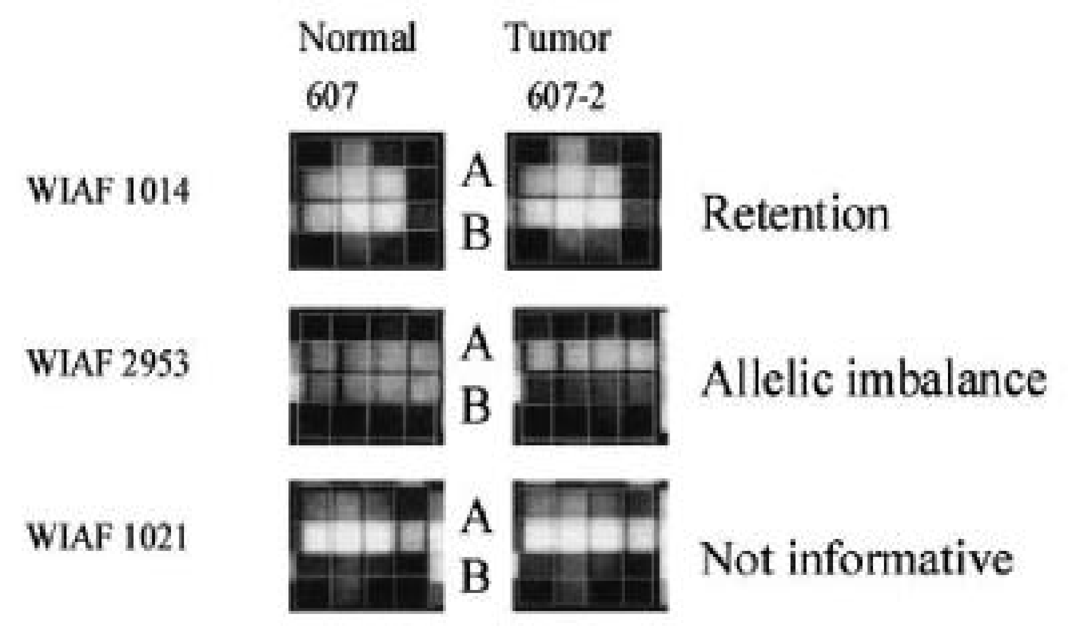
Table 3. ABACUS SNP Detection and Genotyping Accuracy

	Verified	Total Possible
A. Accuracy of autosomal SNPs detection		
Singleton SNPs	17	17
Non-singleton SNPs	91	91
Total SNPs	108	108
B. Number of autosomal SNPs electronically verified		
Number of SNPs electronically verified	371	
C. Accuracy of autosomal genotype calls		
Number of verified homozygous genotype calls	1515	
Number of incorrect homozygous genotype calls	0	
Percent correct homozygote calls	100.00%	
Number of verified heterozygous genotype calls	423	
Number of incorrect heterozygous genotype calls	3	
Percent correct heterozygote calls	99.30%	
D. Accuracy of haploid genotype calls		
Number of bases sequenced (6X coverage)	17,423	
Number of bases different from microarray chip calls	0	
Percent of bases identical	100.00%	

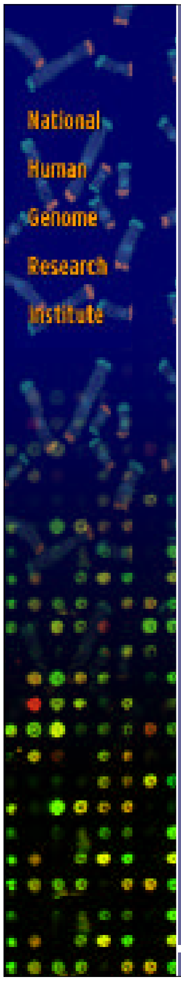
Cutler DJ et al. Genome Res. 2001 11:1913-25.

National
Human
Genome
Research
Institute

SNP CHIP FOR ALLELIC IMBALANCE



Primdahl H et al. J Natl Cancer Inst. 2002, 94:216-223

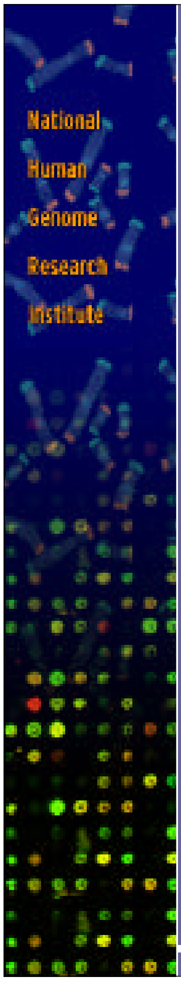


DNA Microarray Applications

- Resequencing

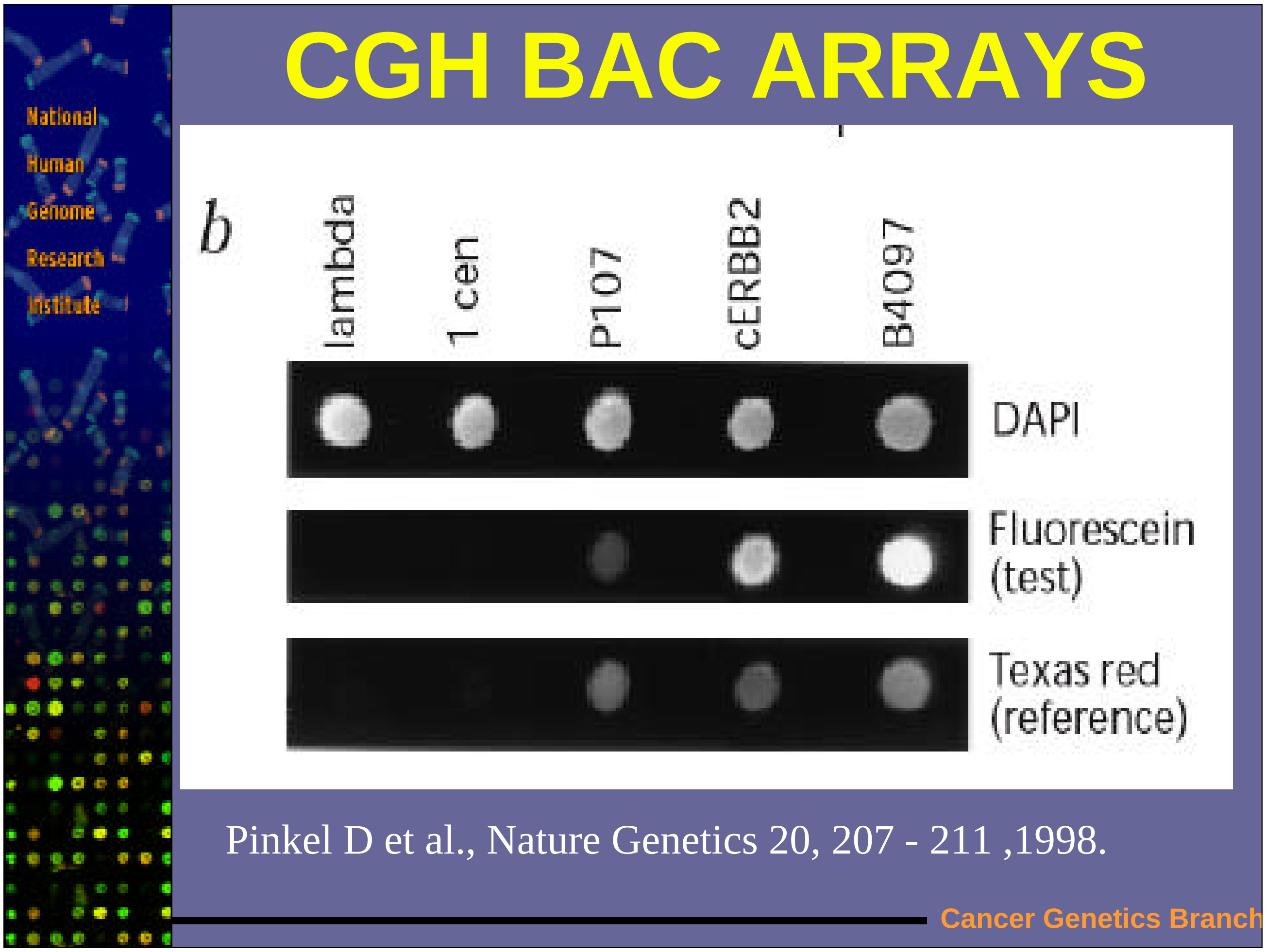
  Mutations
  Polymorphisms

- Gene copy number



Gene Copy Number

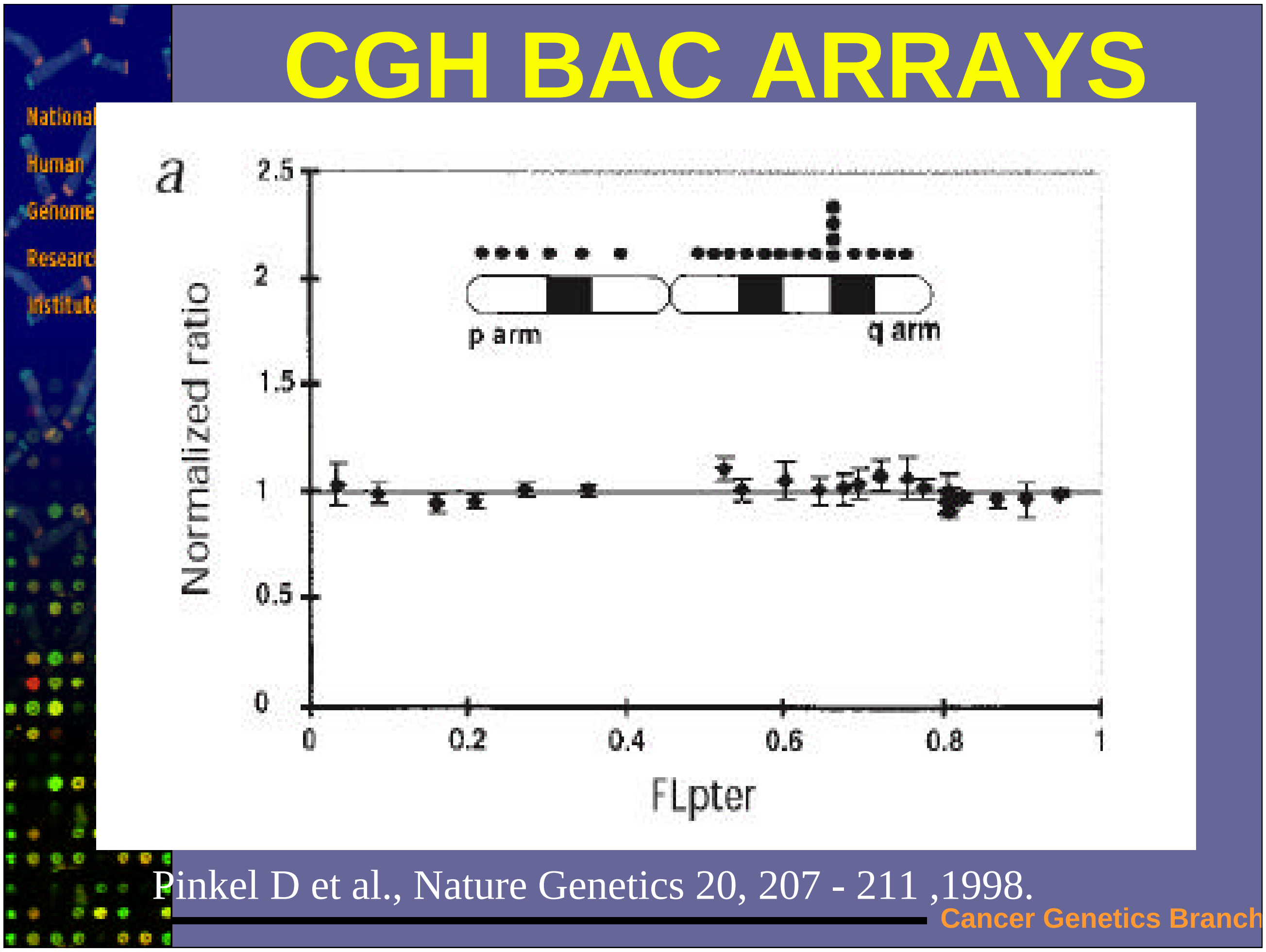
- Array format CGH
- Large insert clones
- cDNA clones/exons



CGH BAC ARRAYS

Pinkel D et al., Nature Genetics 20, 207 - 211 ,1998.

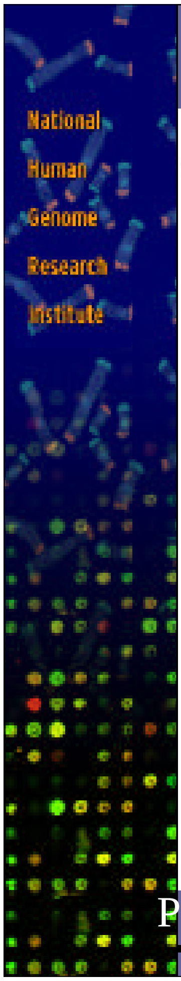
Cancer Genetics Branch



CGH BAC ARRAYS

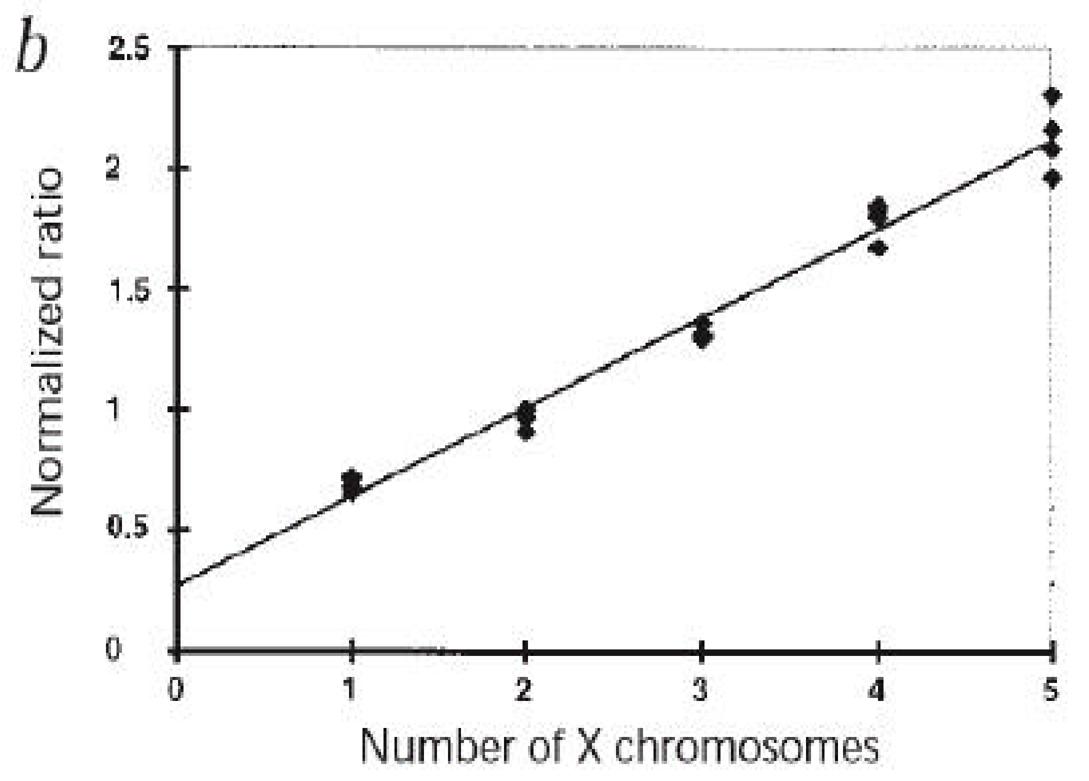
Pinkel D et al., Nature Genetics 20, 207 - 211 ,1998.

Cancer Genetics Branch



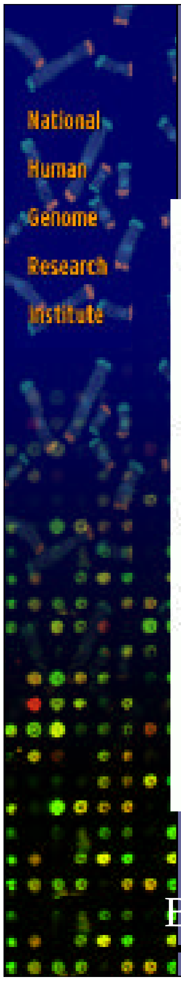
National
Human
Genome
Research
Institute

CGH BAC ARRAYS



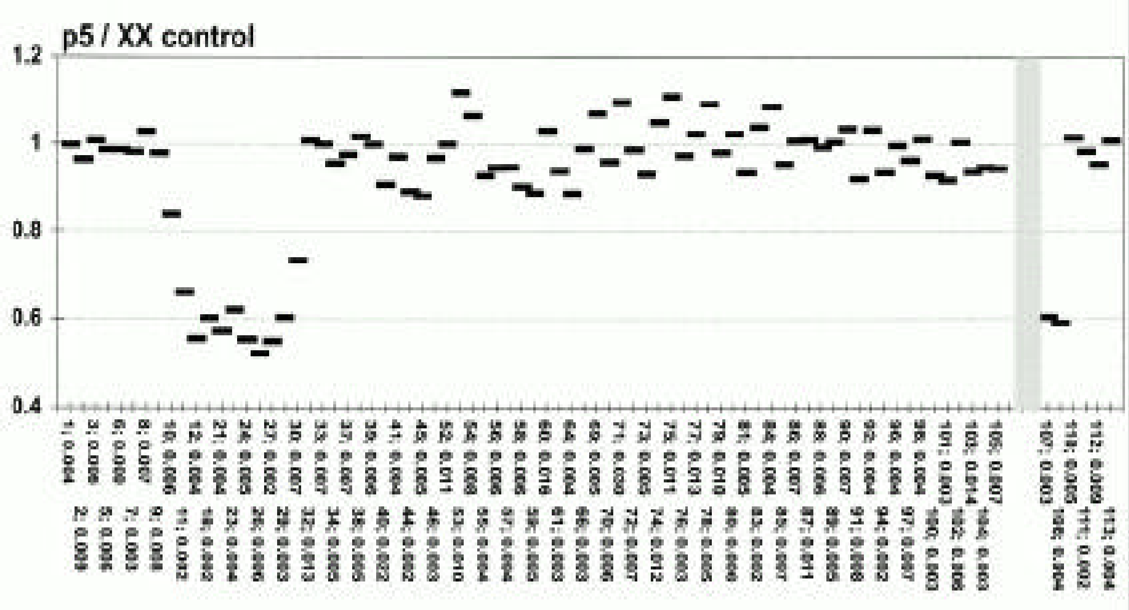
Pinkel D et al., Nature Genetics 20, 207 - 211 ,1998.

Cancer Genetics Branch



National
Human
Genome
Research
Institute

CGH BAC ARRAYS

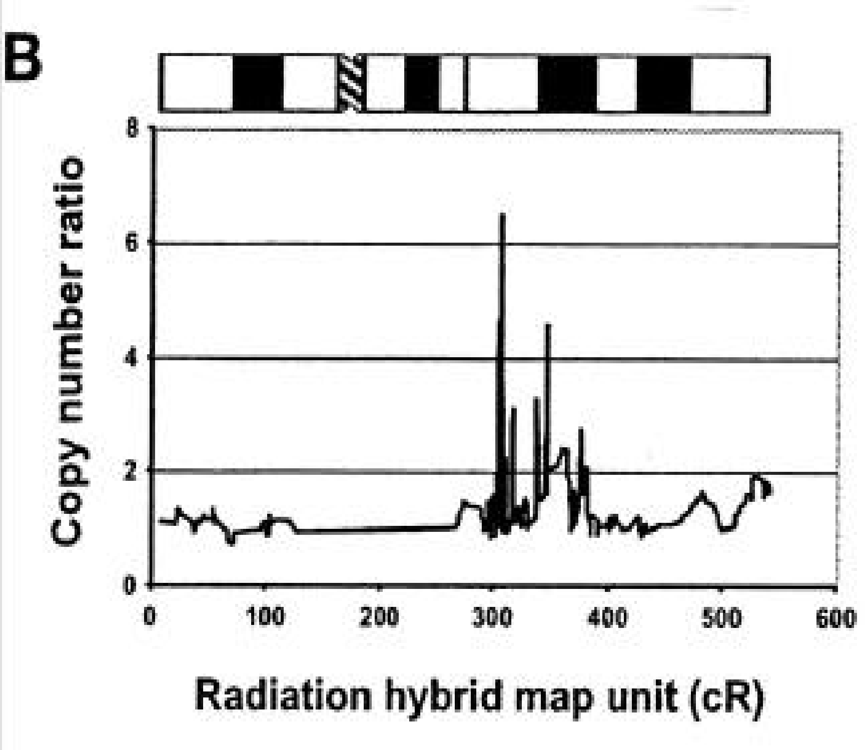


Bruder CE et al., Hum Mol Genet. 2001;10:271-82.

Cancer Genetics Branch

National
Human
Genome
Research
Institute

CGH cDNA

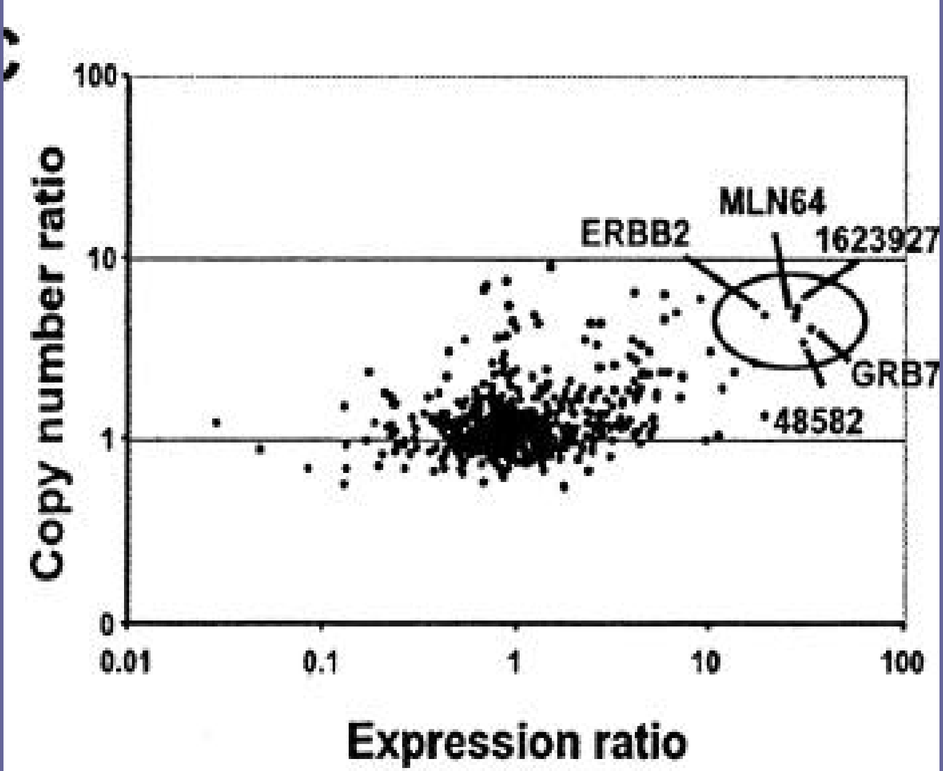


Kauraniemi P et al., Cancer Res. 2001 ;61:8235-40.

Cancer Genetics Branch

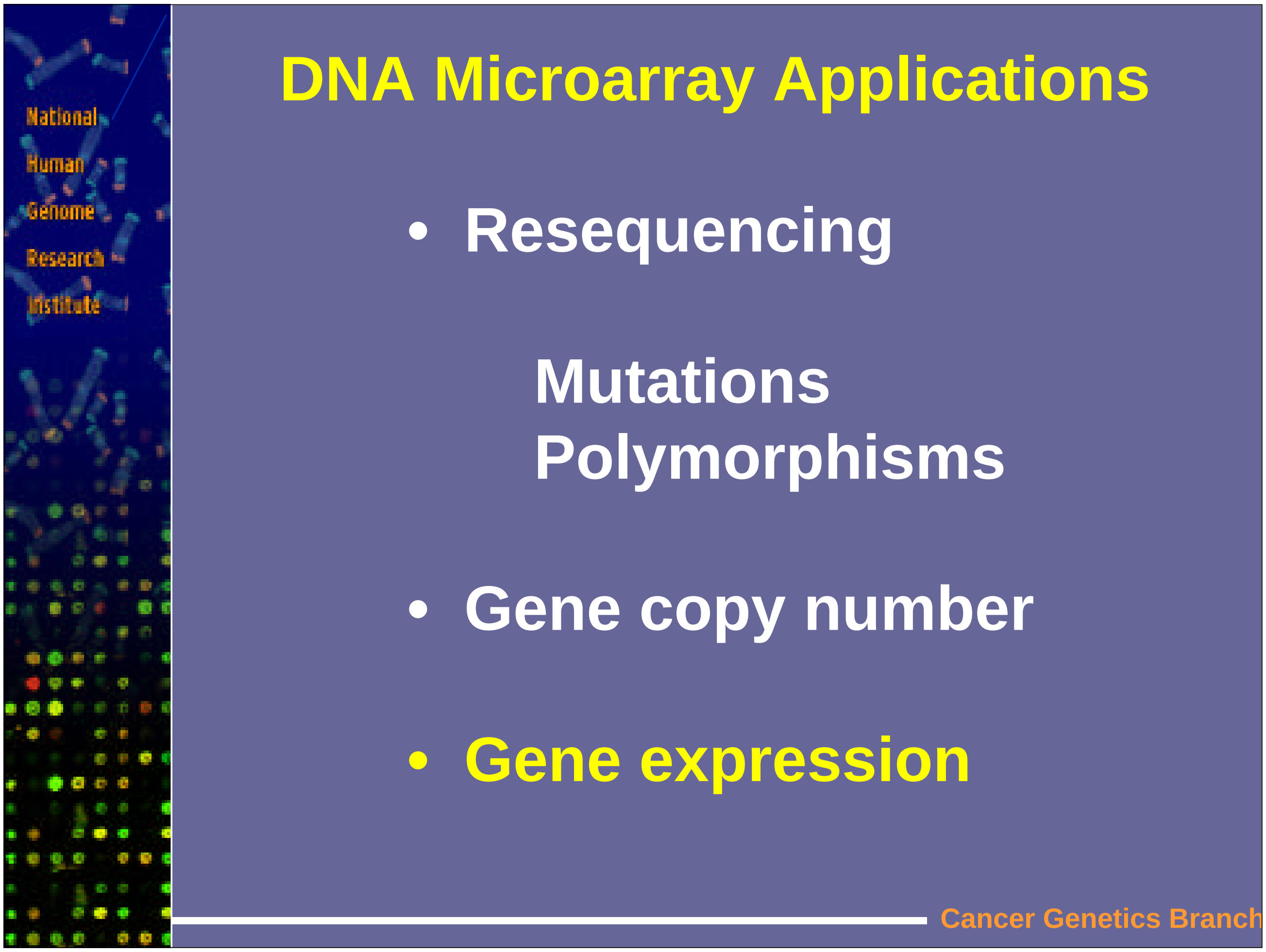
National
Human
Genome
Research
Institute

CGH cDNA



Kauraniemi P et al., Cancer Res. 2001

Cancer Genetics Branch



National
Human
Genome
Research
Institute

DNA Microarray Applications

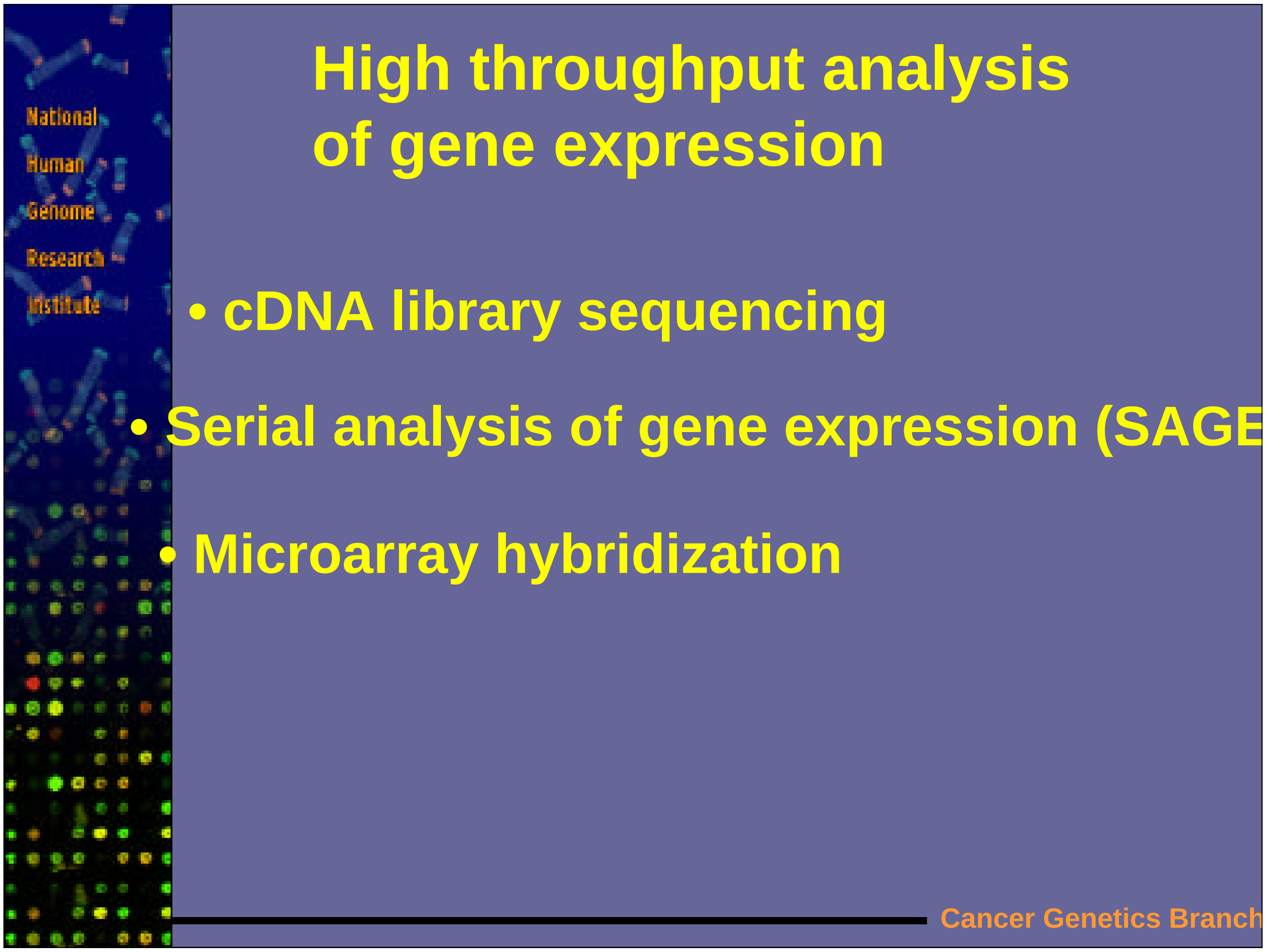
- Resequencing

  Mutations
  Polymorphisms

- Gene copy number

- Gene expression

Cancer Genetics Branch



National
Human
Genome
Research
Institute

High throughput analysis of gene expression

- cDNA library sequencing

- Serial analysis of gene expression (SAGE

- Microarray hybridization

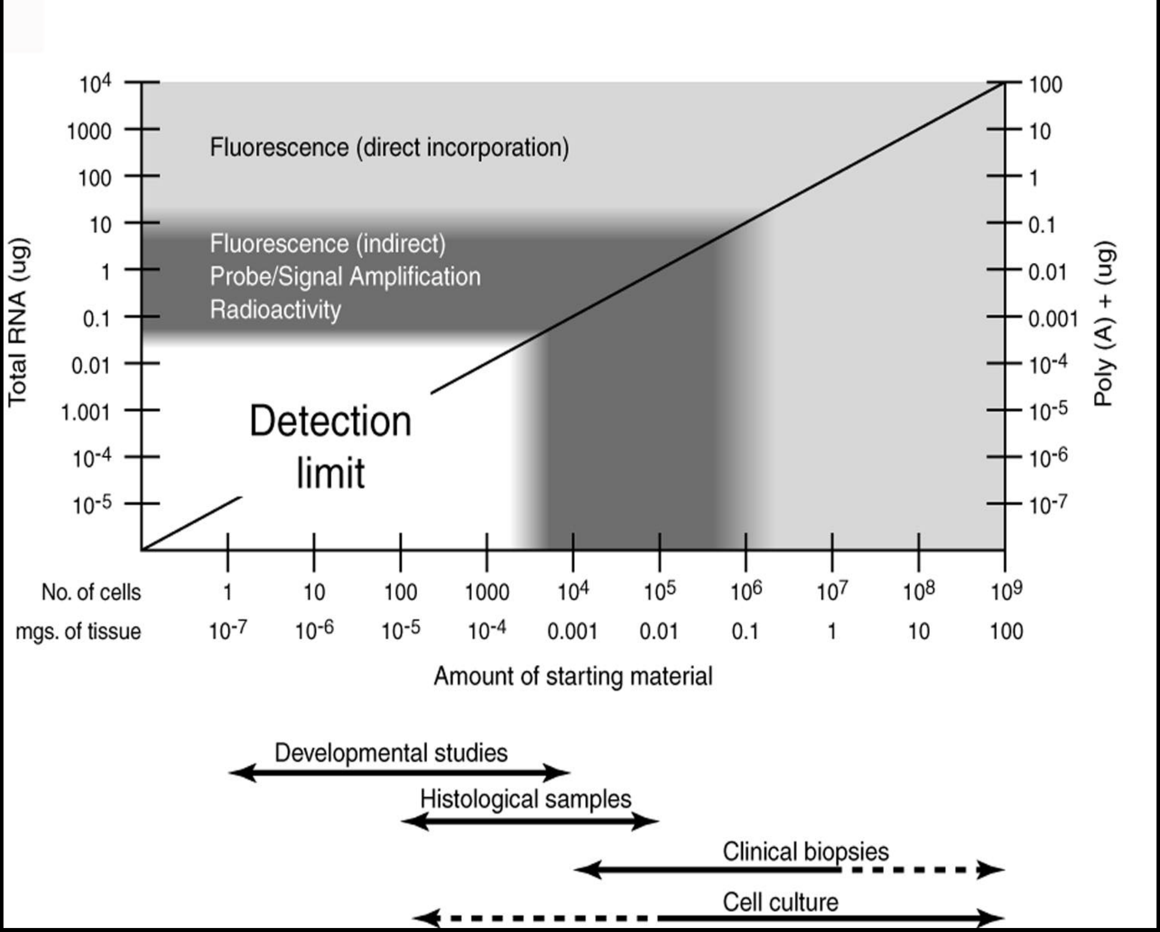
Cancer Genetics Branch

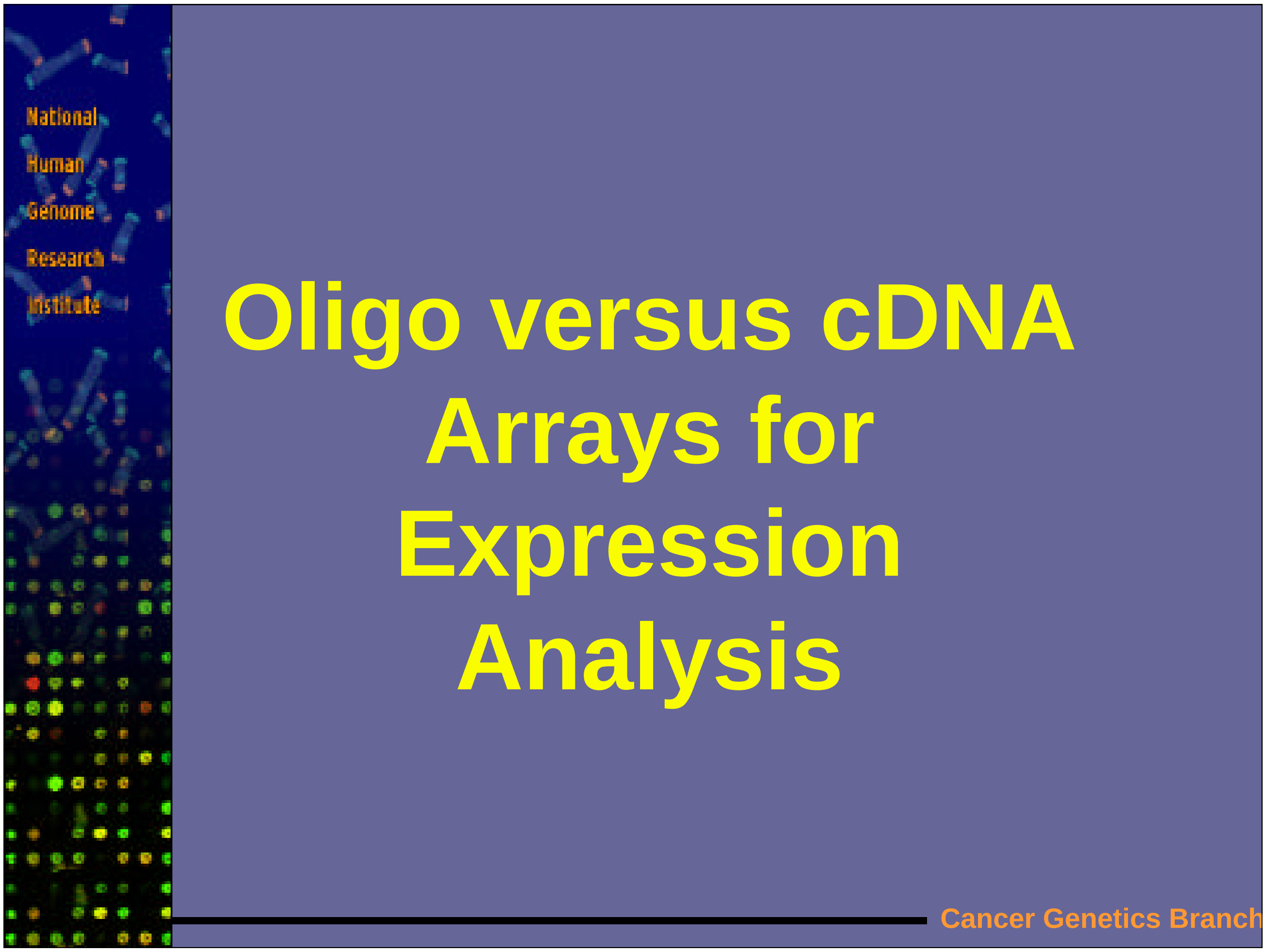
National
Human
Genome
Research
Institute

STRATEGIES FOR SIGNAL GENERATION FROM mRNA

- Fluorochrome conjugated cDNA
- Ligand substituted nucleotides with secondary dete
- Radioactivity
- RNA amplification

Cancer Genetics Branch

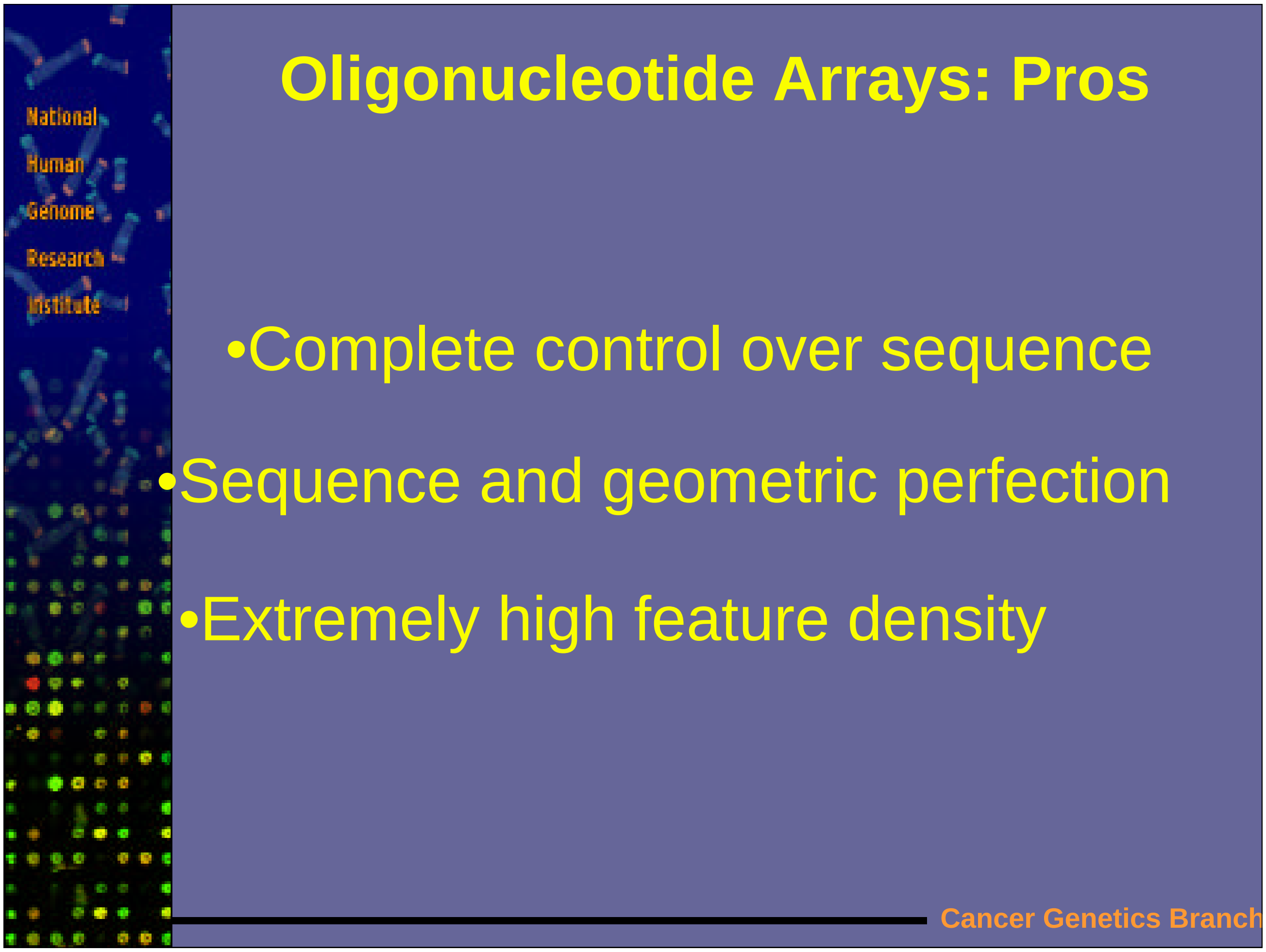




National
Human
Genome
Research
Institute

Oligo versus cDNA Arrays for Expression Analysis

Cancer Genetics Branch

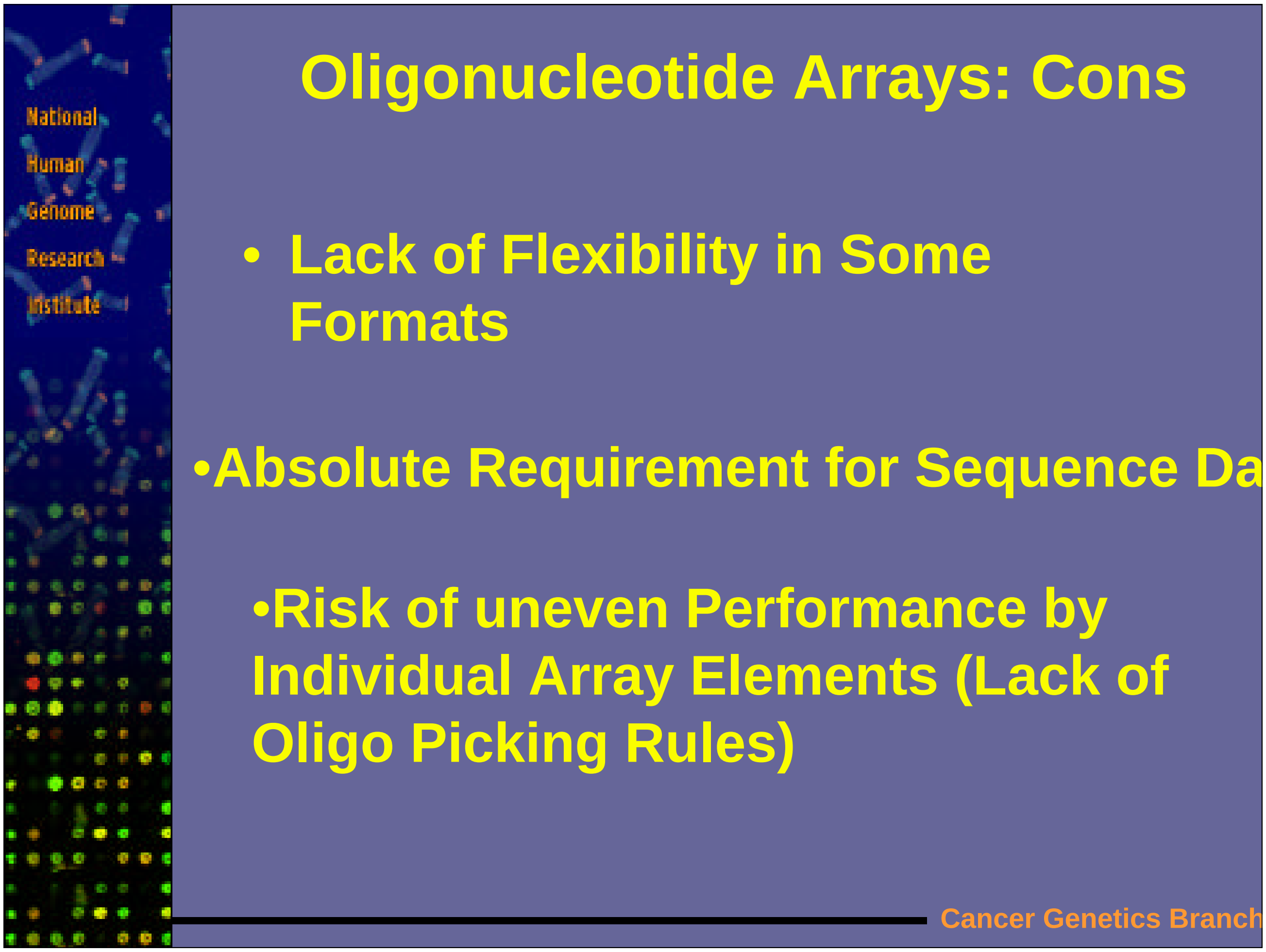


National
Human
Genome
Research
Institute

Oligonucleotide Arrays: Pros

- Complete control over sequence
- Sequence and geometric perfection
- Extremely high feature density

Cancer Genetics Branch



National
Human
Genome
Research
Institute

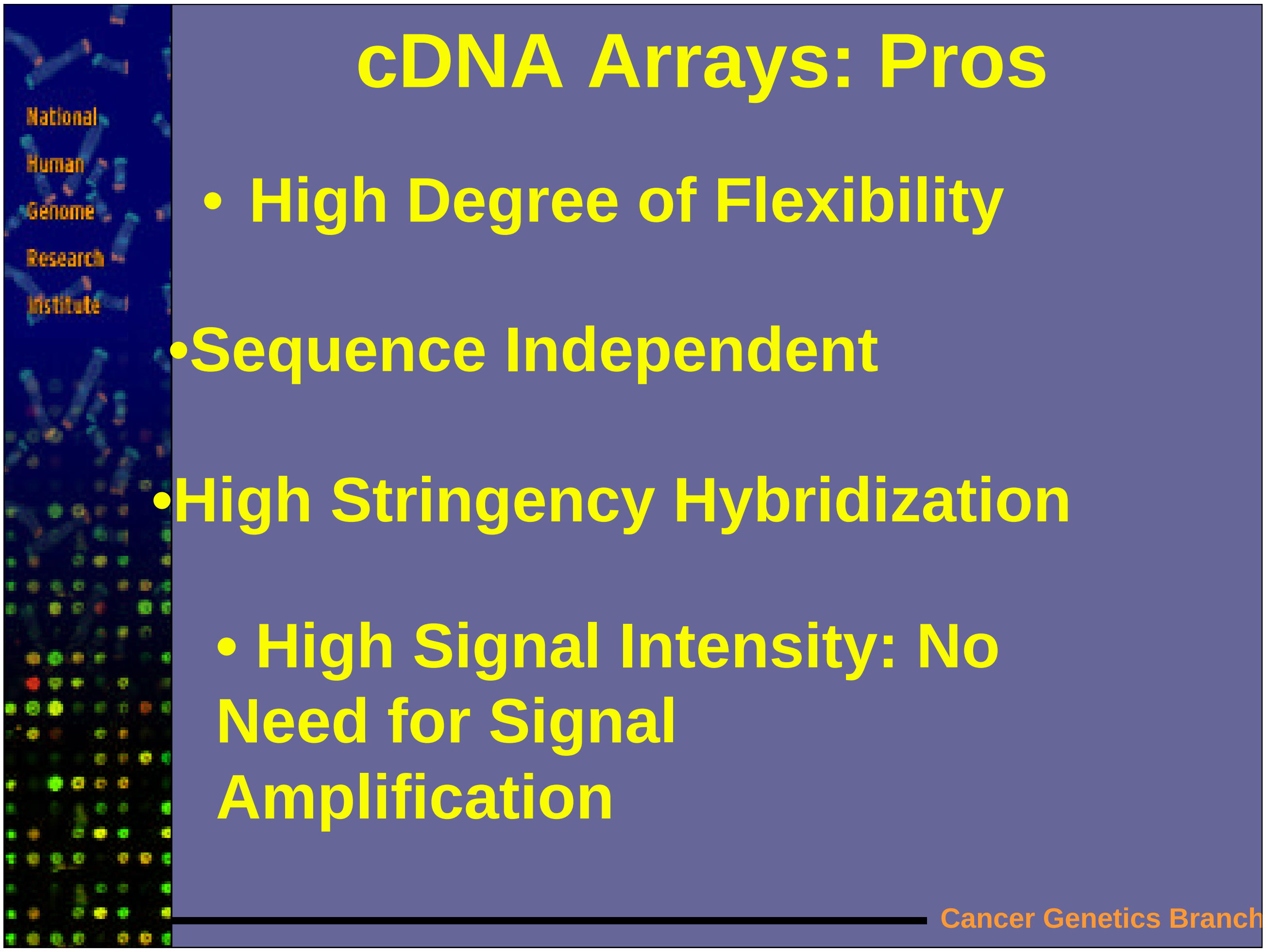
Oligonucleotide Arrays: Cons

- Lack of Flexibility in Some Formats

•Absolute Requirement for Sequence Da

•Risk of uneven Performance by Individual Array Elements (Lack of Oligo Picking Rules)

Cancer Genetics Branch



National
Human
Genome
Research
Institute

cDNA Arrays: Pros

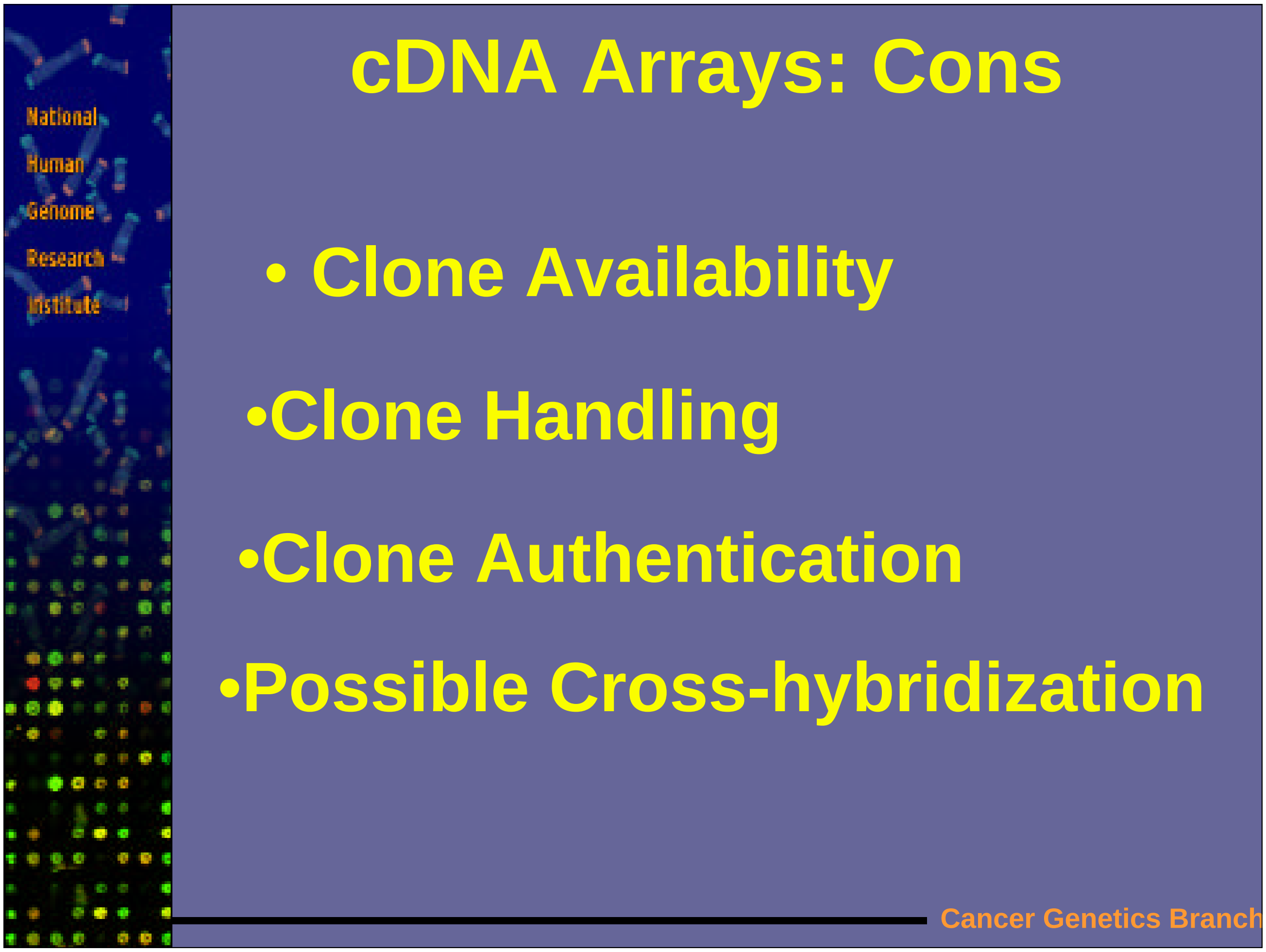
- High Degree of Flexibility

•Sequence Independent

•High Stringency Hybridization

- High Signal Intensity: No Need for Signal Amplification

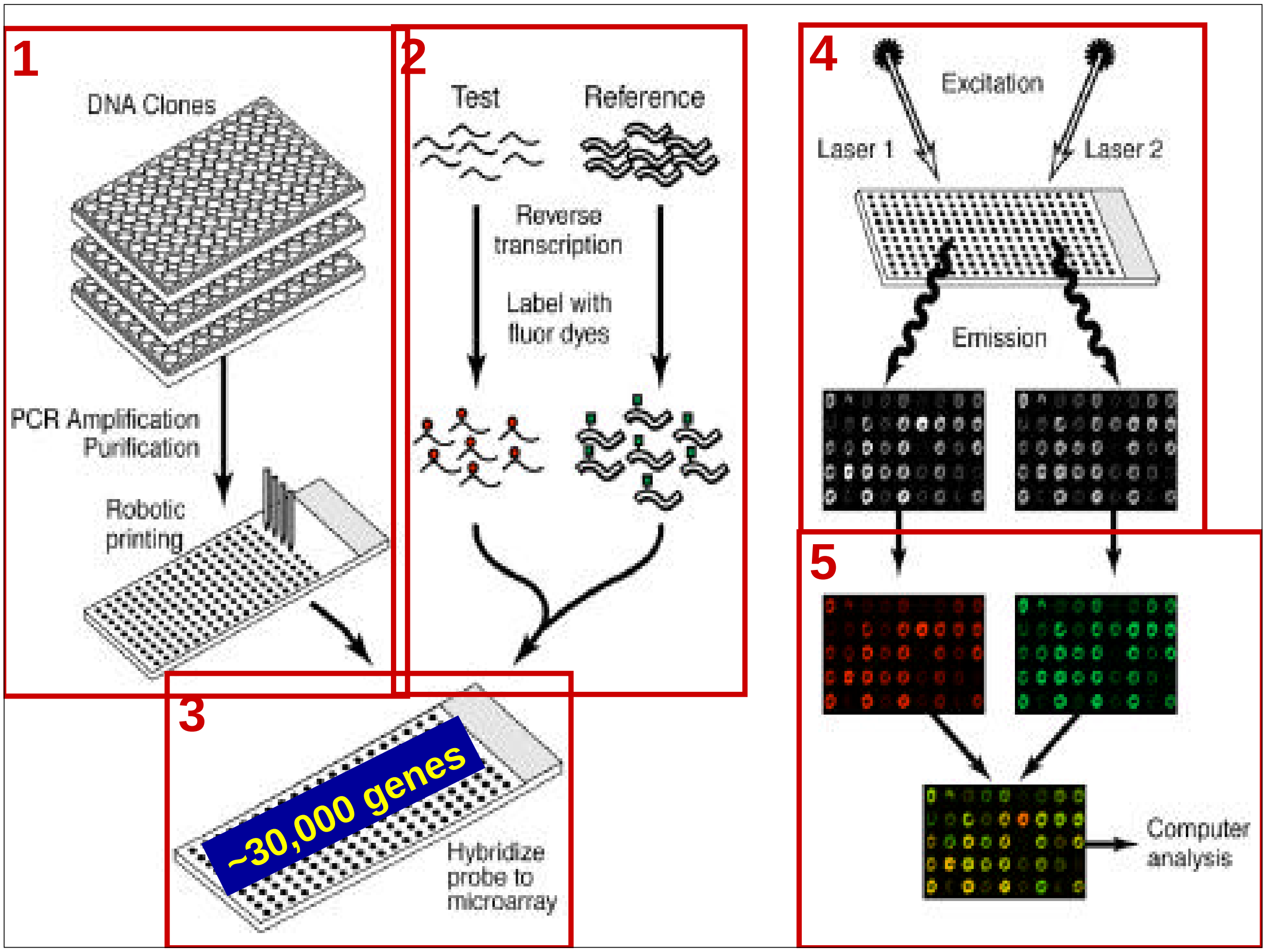
Cancer Genetics Branch


National
Human
Genome
Research
Institute

cDNA Arrays: Cons

- Clone Availability
- Clone Handling
- Clone Authentication
- Possible Cross-hybridization

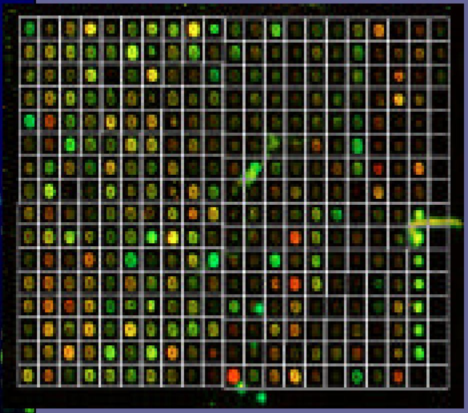
Cancer Genetics Branch



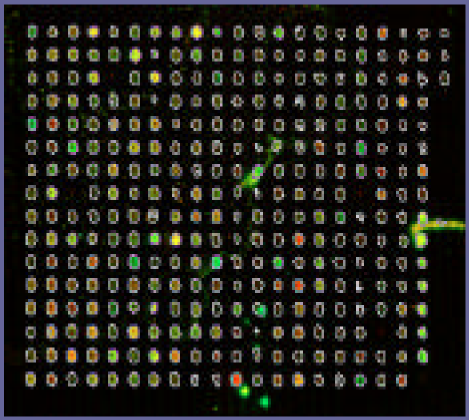
National
Human
Genome
Research
Institute

Image Analysis: DeArray

Grid Overlay



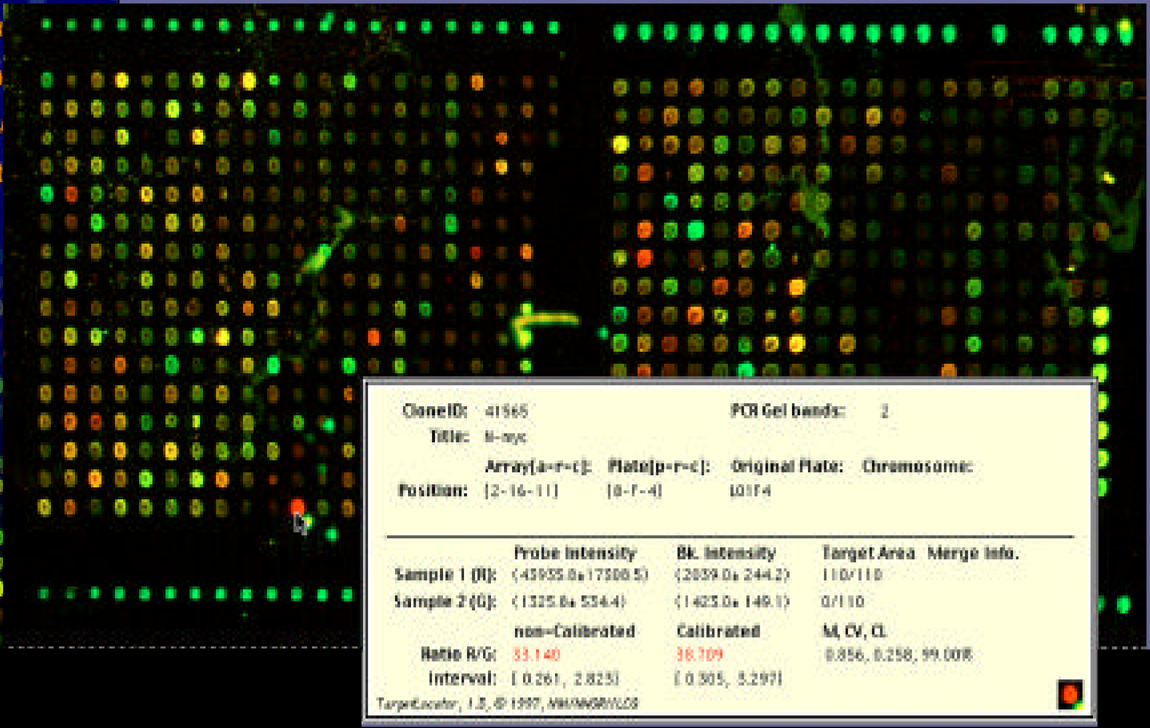
Target detection



Cancer Genetics Branch

National
Hum
Gen
Res
Inst

Image Analysis: DeArray



CloneID: 41565 PCR Gel bands: 2

Title: N-myc

	Array[a-r-c]:	Plate[p-r-c]:	Original Plate:	Chromosome:
Position:	[2-16-11]	[8-F-4]	L01F4	

	Probe Intensity	Bk. Intensity	Target Area	Merge Info.
Sample 1 (R):	(45935.0± 17508.5)	(2039.0± 244.2)	110/110	
Sample 2 (G):	(1325.6± 534.4)	(1425.0± 149.1)	0/110	
	non-Calibrated	Calibrated	M, CV, CL	
Ratio R/G:	33.140	38.709	0.856, 0.258, 99.00%	
Interval:	[0.261 , 2.823]	[0.305, 3.297]		

TargetLocator, 1.5, © 1997, NIH/NHGRI/LCG

DATA QUALITY IS CRITICAL

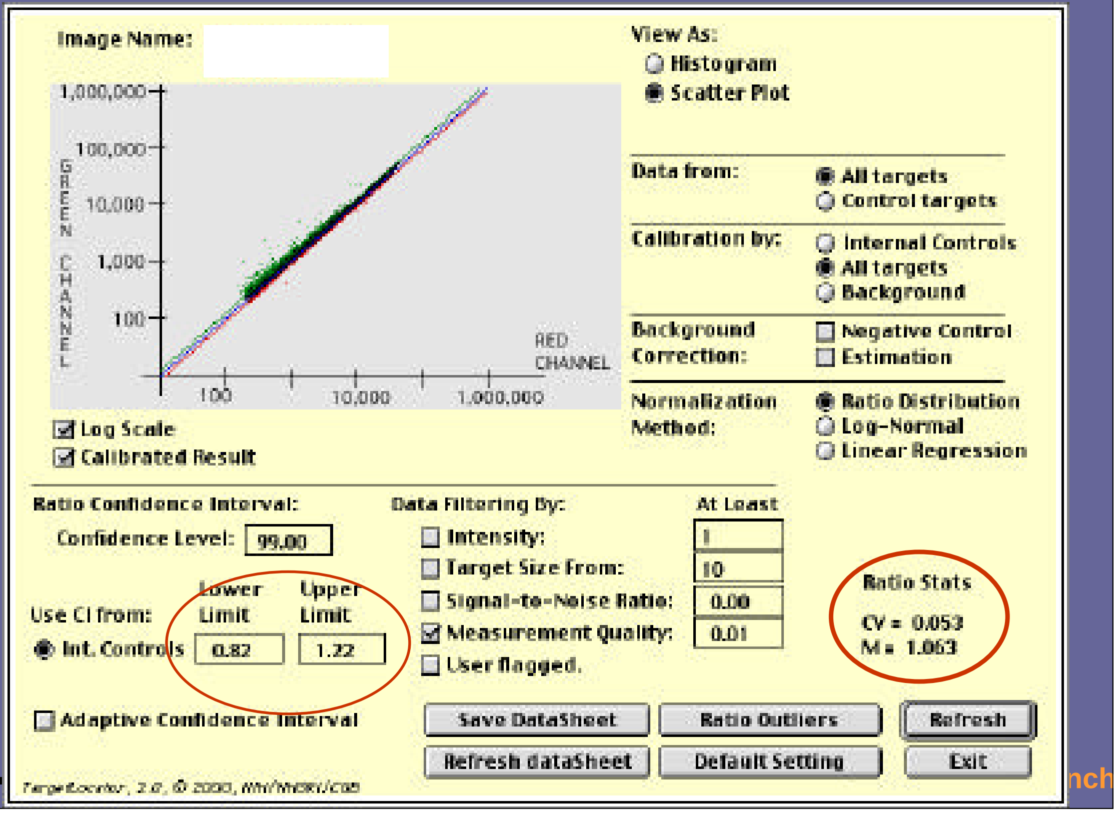


Image Name:

View As:
○ Histogram
● Scatter Plot

☑ Log Scale
☑ Calibrated Result

Data from: ● All targets ○ Control targets

Calibration by: ○ Internal Controls ● All targets ○ Background

Background Correction: □ Negative Control □ Estimation

Normalization Method: ● Ratio Distribution ○ Log-Normal ○ Linear Regression

Ratio Confidence Interval:
Confidence Level: 99.00

Use CI from:

	Lower Limit	Upper Limit
● Int. Controls	0.82	1.22

□ Adaptive Confidence Interval

Data Filtering By:

	At Least
□ Intensity:	1
□ Target Size From:	10
□ Signal-to-Noise Ratio:	0.00
☑ Measurement Quality:	0.01
□ User flagged.	

Ratio Stats
CV = 0.053
M = 1.063

TargetLocator, 2.0, © 2000, NIH/NHGRI/CGB

Output of cDNA microarrays: expression ratio

Output of oligonucleotide arrays: expression lev

Both types of data can be analyzed with essentially the same tools.

National
Human
Genome
Research
Institute

Normalization of cDNA microarrays:

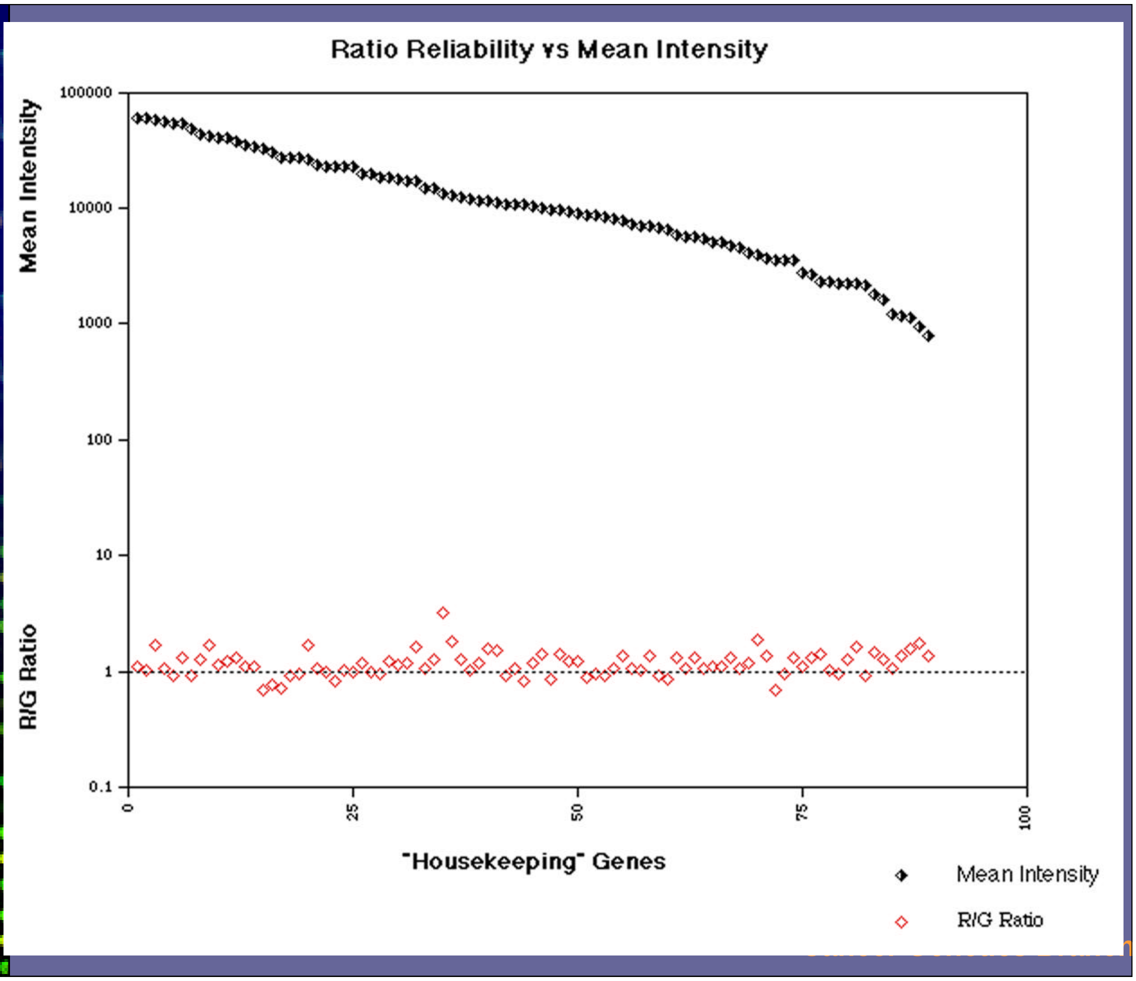
- global
- housekeeping
- spiked standards

Normalization of oligonucleotide arrays:

- global

Cancer Genetics Branch

National
Human
Genome
Research
Institute

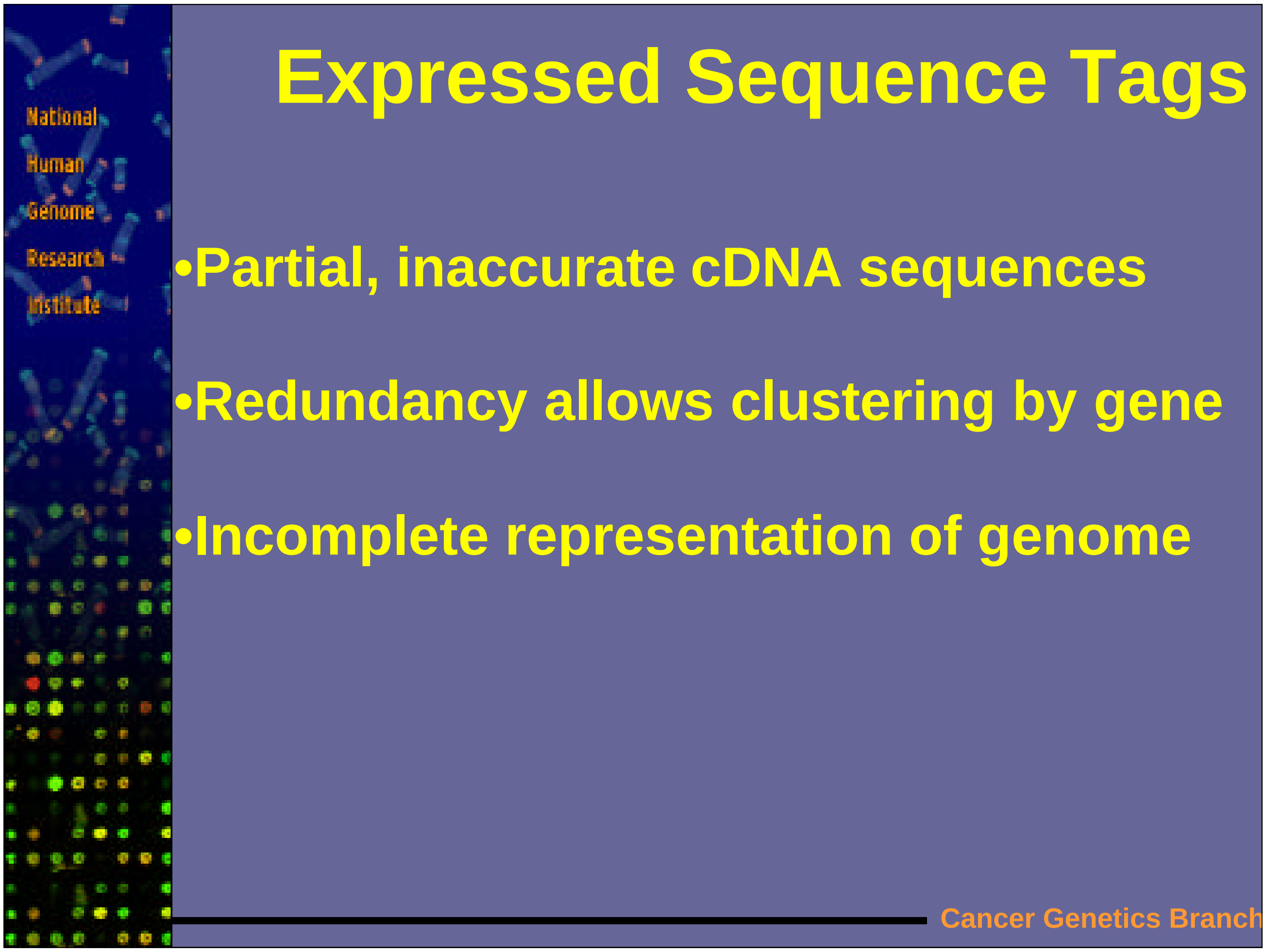


National
Human
Genome
Research
Institute

Building Expression Arrays

- Completely sequenced and annotated organisms
- Complete gene list unavailable

Cancer Genetics Branch


National
Human
Genome
Research
Institute

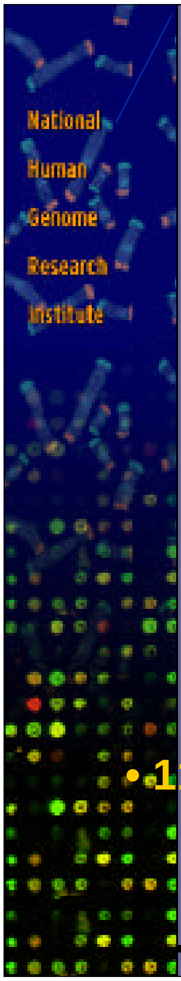
Expressed Sequence Tags

•Partial, inaccurate cDNA sequences

•Redundancy allows clustering by gene

•Incomplete representation of genome

Cancer Genetics Branch



Clustering of Human ESTs

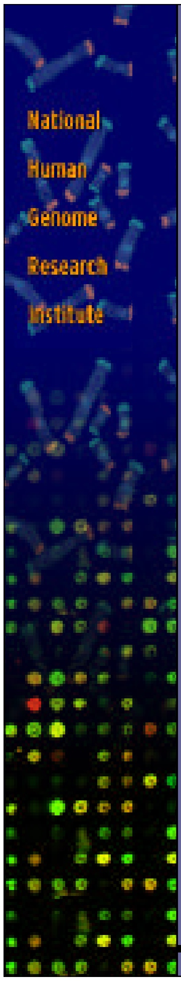
Final Number of Clusters (sets) (Unigene 146)
==

96109 sets total

21857 sets contain at least one known gene
94916 sets contain at least one EST
20664 sets contain both genes and ESTs

- **1193 genes are not represented by an identifiable EST.**

Cancer Genetics Branch



Histogram of cluster sizes for UniGene build 146

96,000 clusters

Cluster size	Number of clusters
1	33700
2	13467
3-4	15267
5-8	10197
9-16	5777
17-32	3894
33-64	3549
65-128	4031
129-256	3718
257-512	1732
513-1024	537
1025-2048	162
2049-4096	56
4097-8192	19
8193-16384	3

Cancer Genetics Branch

National
Human
Genome
Research
Institute

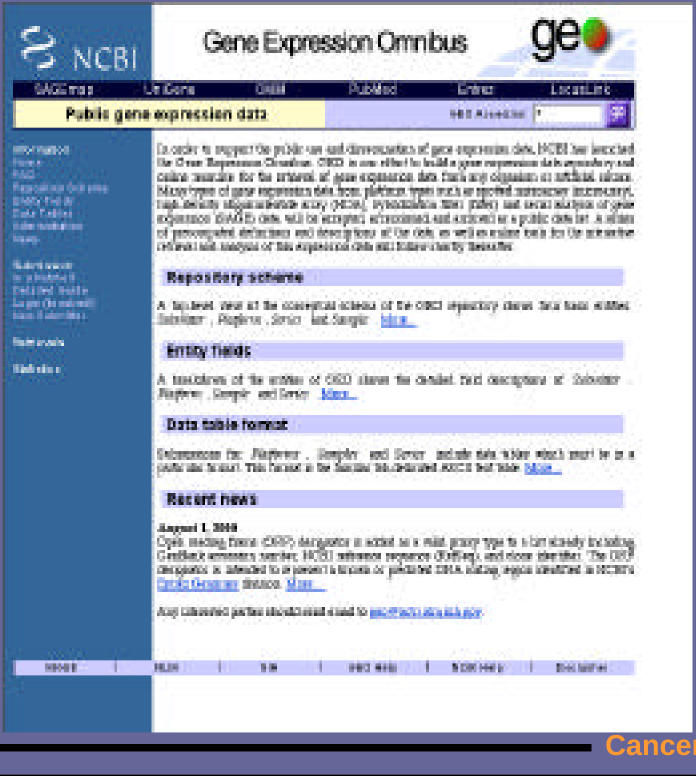
Expressed Sequence Tags: Options for Array Constructi

- ”Standard” clone sets
- Custom clone sets
- Synthetic oligonucleotides

National
Human
Genome
Research
Institute

Accessing Expression Data

- Individual Lab and Journal Sites



In order to support the public use and dissemination of gene expression data, NCBI has launched the Gene Expression Omnibus, GEO [illegible]

Repository scheme

A top-level view of the conceptual scheme of the GEO repository [illegible] Submitter, Platform, Series and Sample. More...

Entity fields

A breakdown of the entities of GEO [illegible] Submitter, Platform, Sample and Series. More...

Data table format

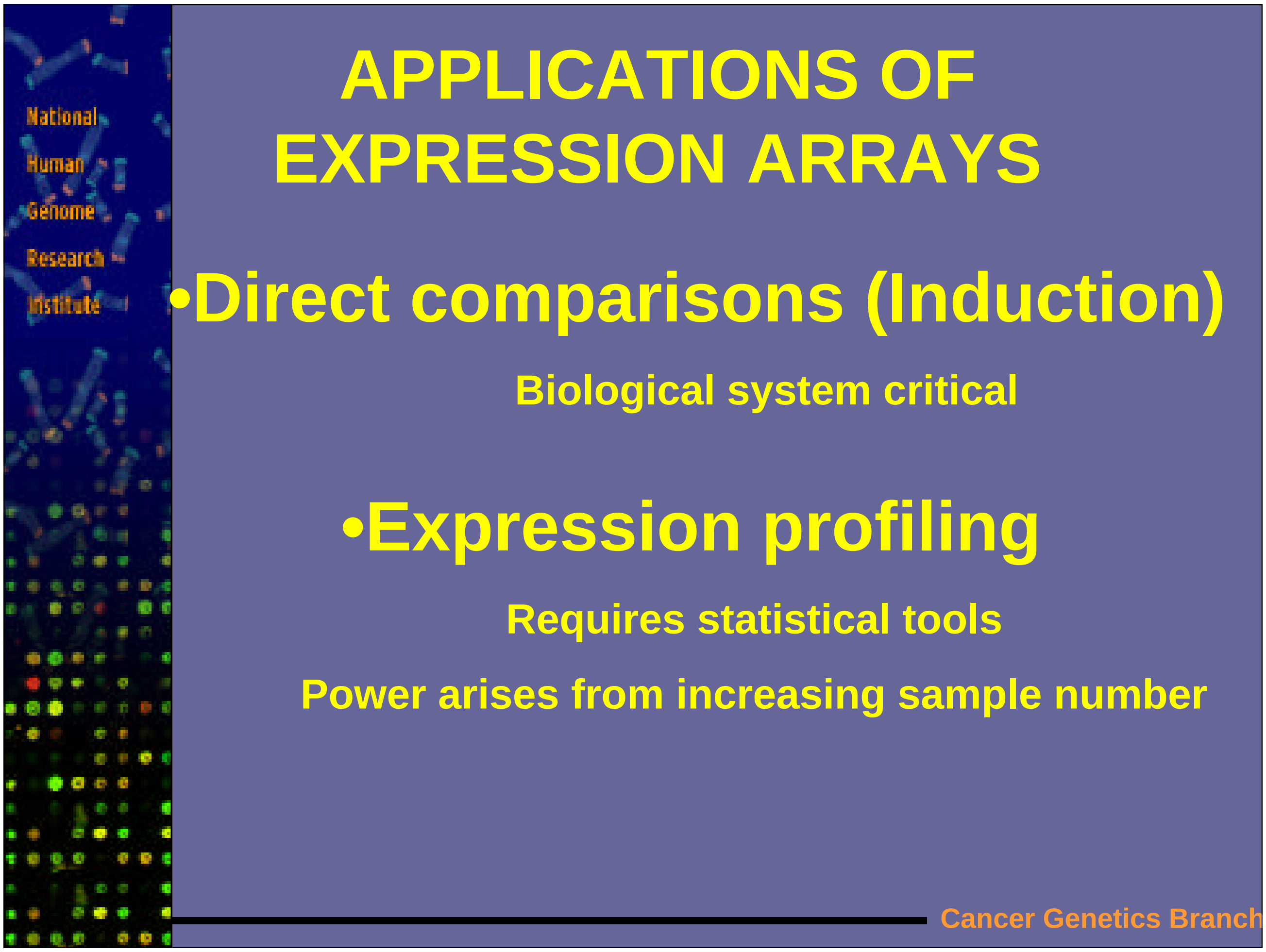
Submissions for Platforms, Samples and Series [illegible] More...

Recent news

August 1, 2002

Open reading frame (ORF) designation [illegible] GenBank [illegible] NCBI [illegible] (RefSeq), and clone identifier. The ORF designation [illegible] Genome [illegible] More...

Any [illegible]



National Human Genome Research Institute

APPLICATIONS OF EXPRESSION ARRAYS

•Direct comparisons (Induction)

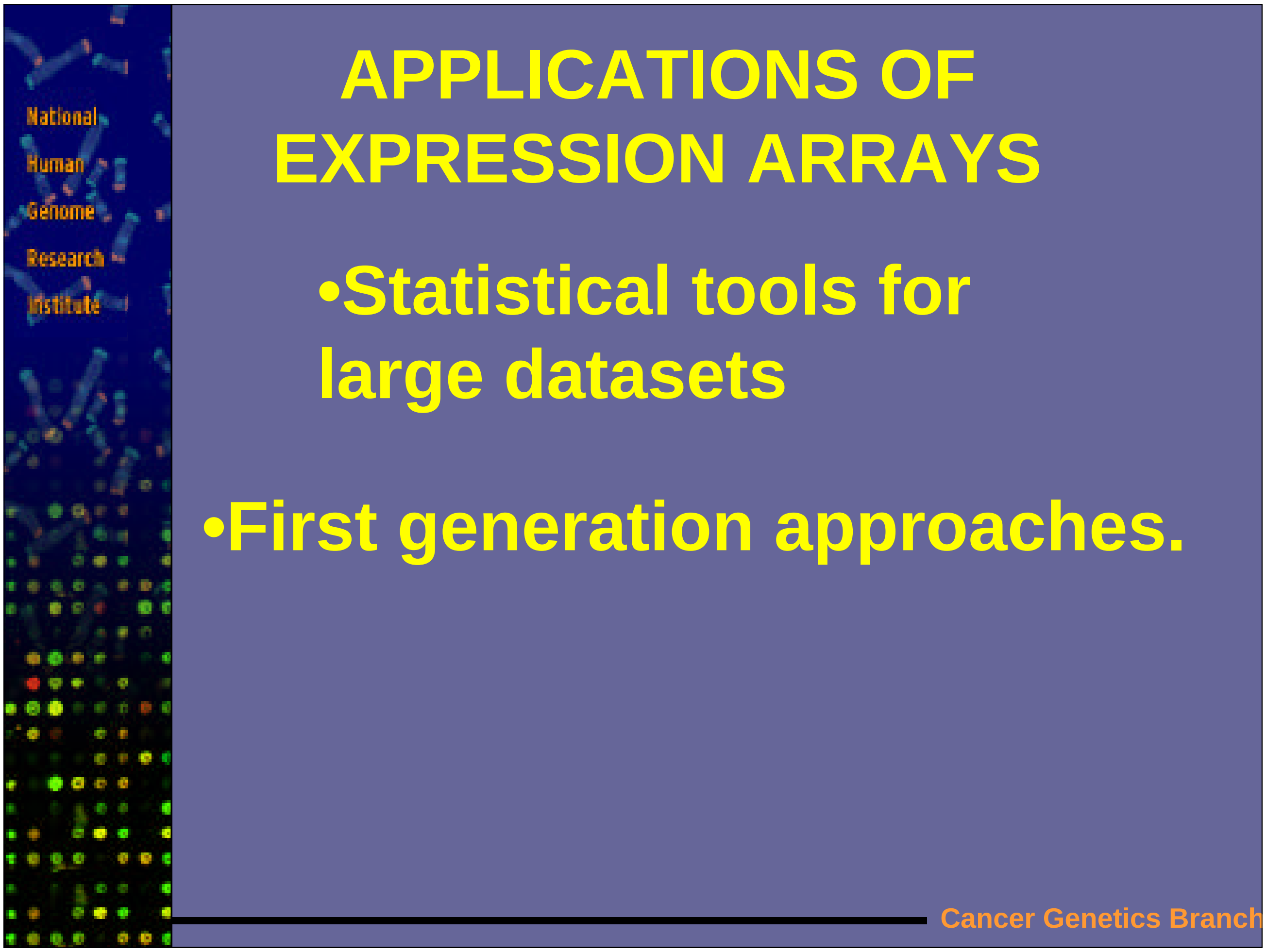
Biological system critical

•Expression profiling

Requires statistical tools

Power arises from increasing sample number

Cancer Genetics Branch



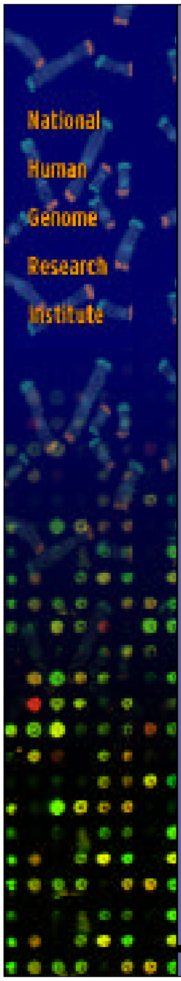
National Human Genome Research Institute

APPLICATIONS OF EXPRESSION ARRAYS

•Statistical tools for large datasets

•First generation approaches.

Cancer Genetics Branch



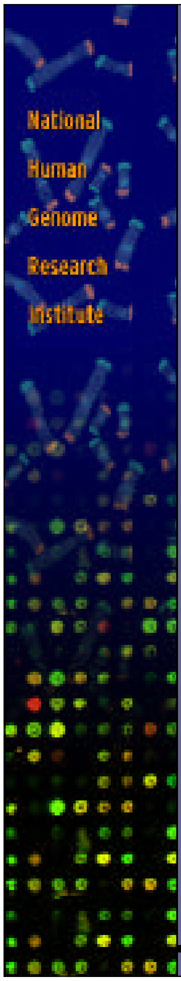
APPLICATIONS OF EXPRESSION ARRAYS : TUMOR PROFILING

- Towards a molecular taxonomy of cancer

•Methods lead to gene identification

•Individualized diagnosis and therapy

Cancer Genetics Branch



APPLICATIONS OF EXPRESSIONARRAYS : GENE IDENTIFICATION

•Groups of genes

•Pathways

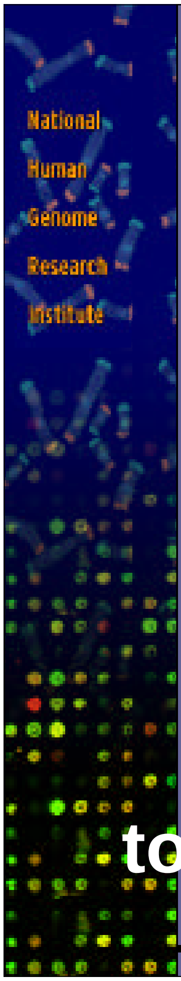
•Co-regulated

•Correlate with copy #

•Correlate clinically

- Candidate disease genes

Cancer Genetics Branch



APPLICATIONS OF EXPRESSION ARRAYS: TUMOR PROFILING

- •Clustering
  - Unsupervised
  - Supervised
- •Classification

•Can classify with respect to any clinically interesting variabl

Cancer Genetics Branch

Alveolar Rhabdomyosarcoma

Pax3
chromosome 2

Method

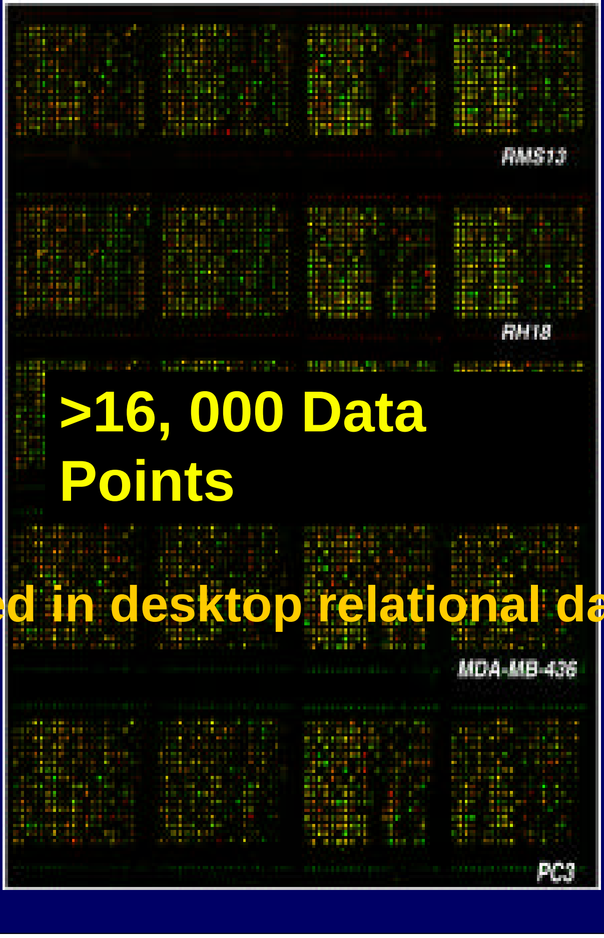
- Compared 7 ARMS with 6 unrelated cancers cell lines
- Using cDNA microarray containing 1238 elements

Cell Line Characteristics

Cell Line	Pax3-FKHR	Diagnosis
ARMS1	+	ARMS
RH3	+	ARMS
RH4	+	ARMS
RH5	+	ARMS
RMS18	+	ARMS
RMS13	+	ARMS
RH28	+	ARMS
A204	-	Undifferentiated Sarcoma
NGP127	-	Neuroblastoma
TC71	-	Ewing's Sarcoma
UACC-903	-	Melanoma
PC3	-	Prostate Carcinoma
MDA-MB-436	-	Breast Carcinoma
Control		
NIL-C	-	Fibroblast

Results

- 13 Experiments
- 1238 genes



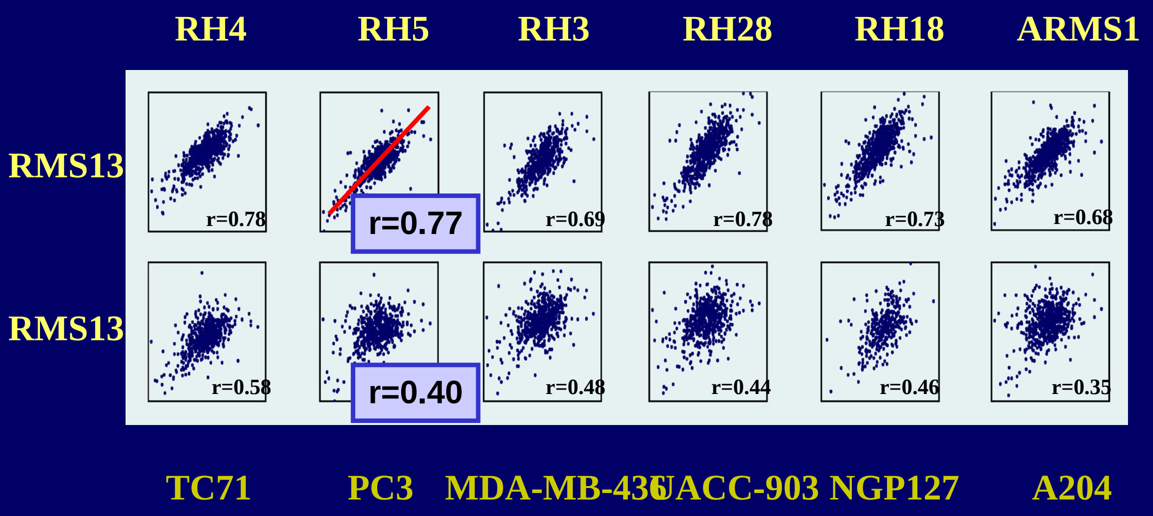
>16, 000 Data Points

Data placed in desktop relational database

Raw Data Matrix

SCATTER PLOT

ARMS

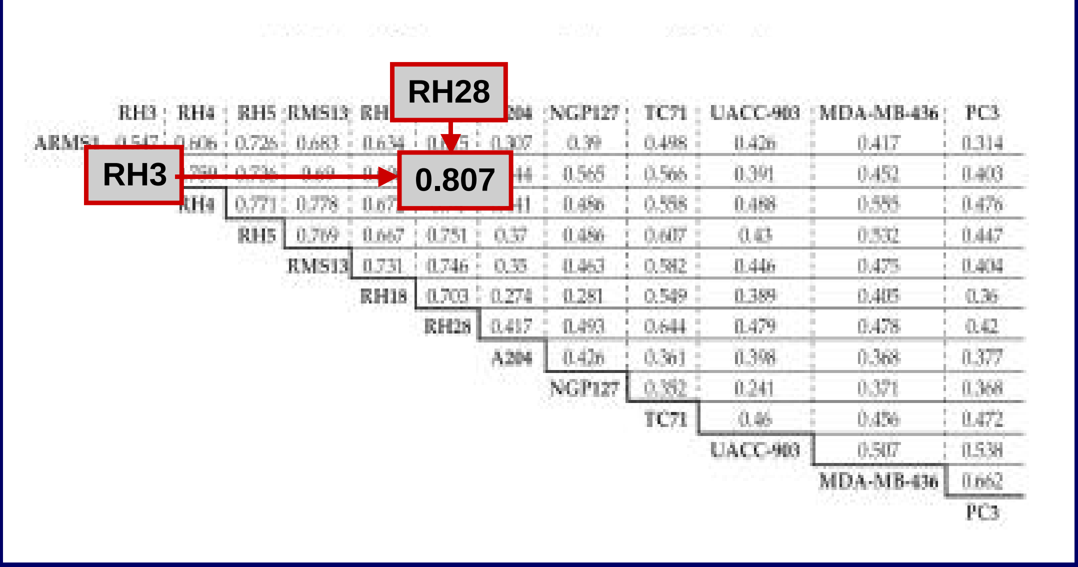


NON-ARMS

Raw Data Matrix

Matrix of Pearson Correlation Coefficients Distance Map

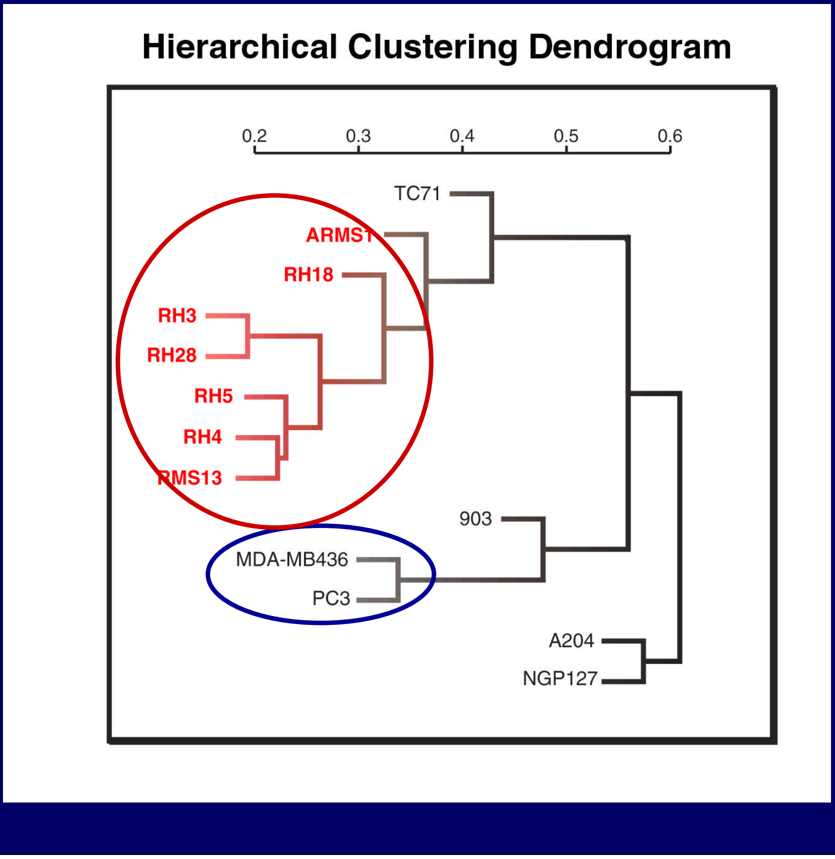
78 pair-wise comparisons



RH28
RH3
0.807

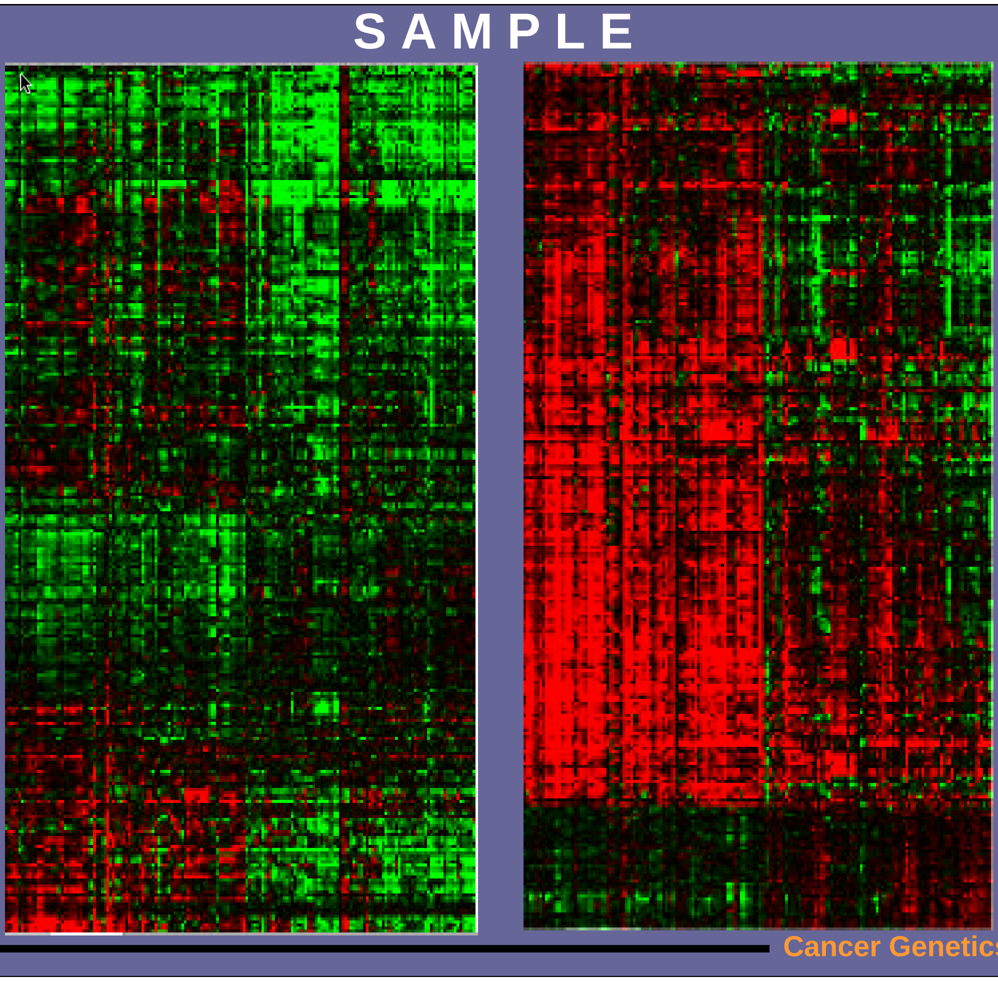
	RH3	RH4	RH5	RMS13	RH18	RH28	A204	NGP127	TC71	UACC-903	MDA-MB-436	PC3
ARMS1	0.547	0.606	0.726	0.683	0.634	[illegible]	0.307	0.39	0.498	0.426	0.417	0.314
RH3		[illegible]	0.736	0.69	[illegible]	0.807	[illegible]	0.565	0.566	0.391	0.452	0.403
RH4			0.771	0.778	[illegible]	[illegible]	[illegible]	0.496	0.558	0.488	0.555	0.476
RH5				0.769	0.667	0.751	0.37	0.486	0.607	0.43	0.532	0.447
RMS13					0.731	0.746	0.35	0.463	0.582	0.446	0.475	0.404
RH18						0.703	0.274	0.281	0.549	0.389	0.405	0.36
RH28							0.417	0.493	0.644	0.479	0.478	0.42
A204								0.426	0.361	0.398	0.368	0.377
NGP127									0.352	0.241	0.371	0.368
TC71										0.46	0.456	0.472
UACC-903											0.507	0.538
MDA-MB-436												0.662
PC3												

Hierarchical Clustering Dendrogram



National Human Genome Research Institute

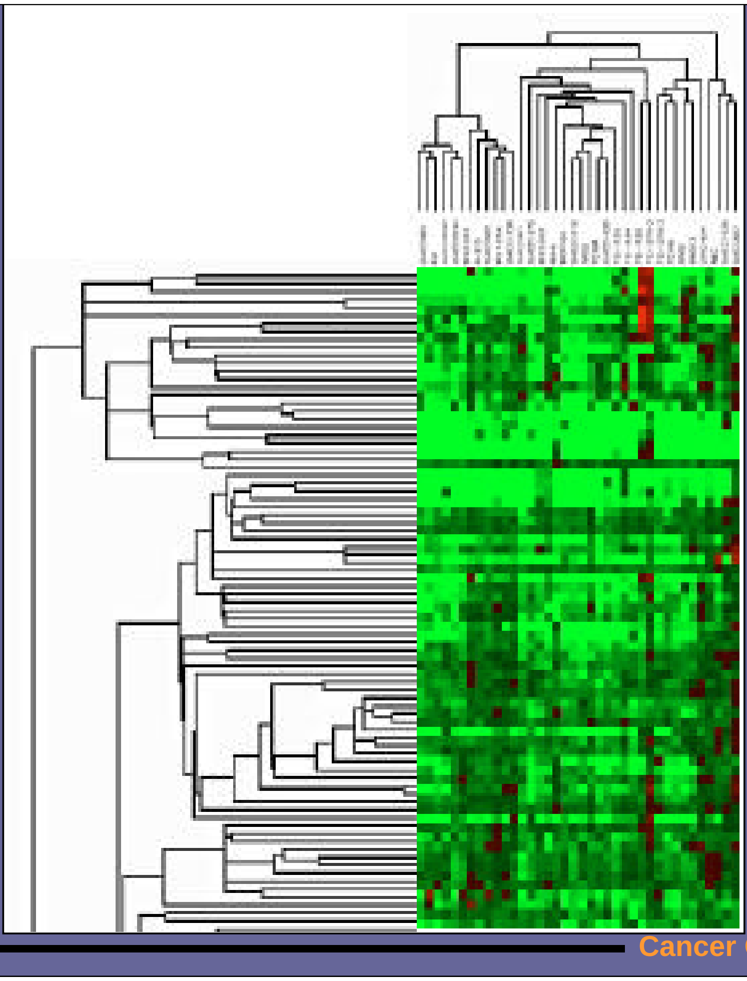
SAMPLE



GENES

Cancer Genetics Branch

National Human Genome Research Institute



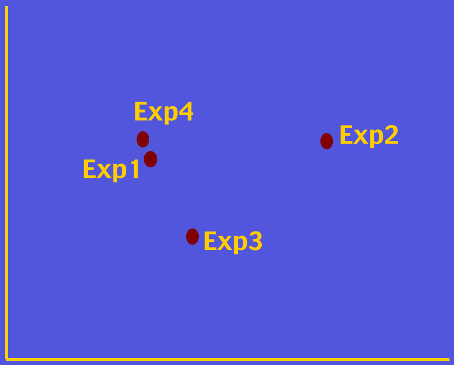
Cancer Genetics Branch

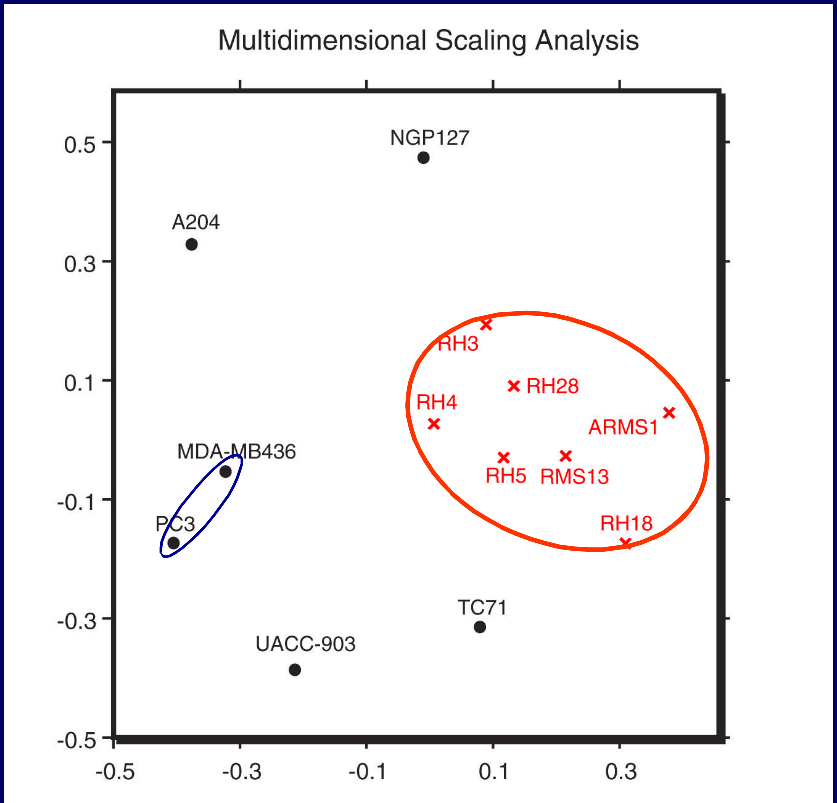
Multidimensional Scaling

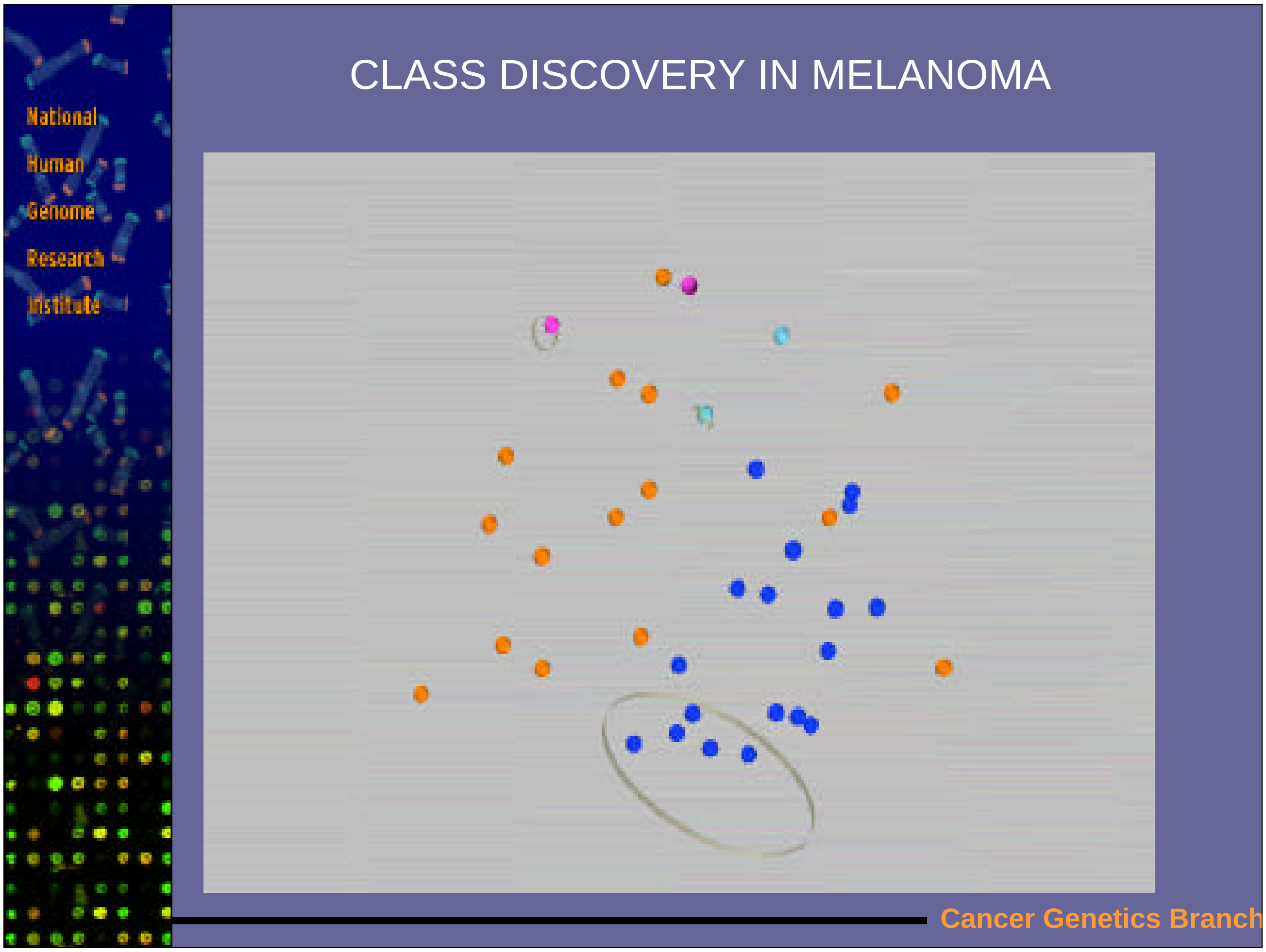
Similarity Measure
Distance Measure

	Seattle	L.A.	Houston	D.C.
Seattle		0	1,200	2,500 2,800
L.A.		0	1,500	2,600
Houston			0	1400
D.C.				0

	Exp1	Exp2	Exp3	Exp4
Exp1	0	0.50	0.25	0.05
Exp2		0	0.50	0.55
Exp3			0	0.60
Exp4				0



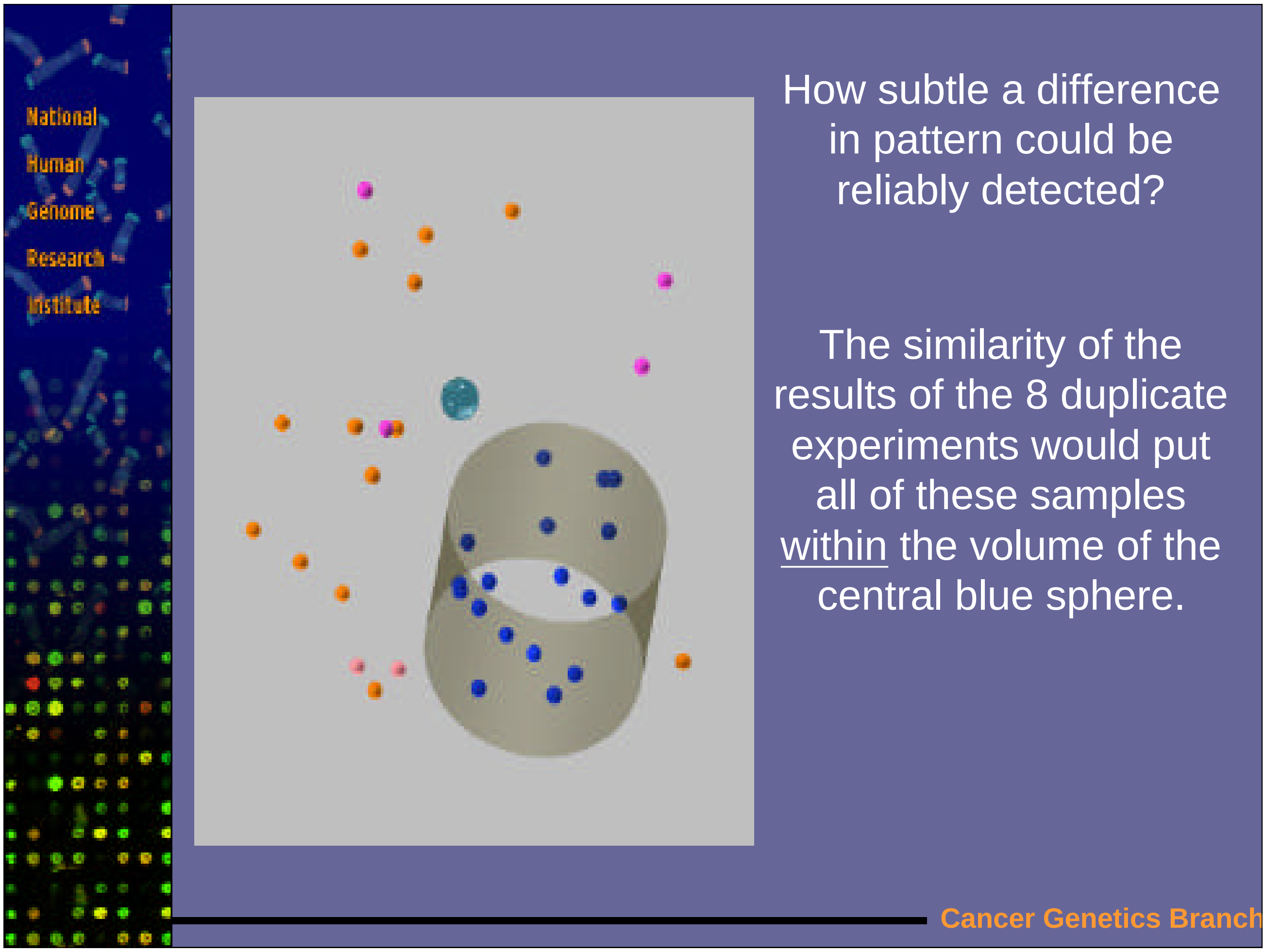




National
Human
Genome
Research
Institute

CLASS DISCOVERY IN MELANOMA

Cancer Genetics Branch



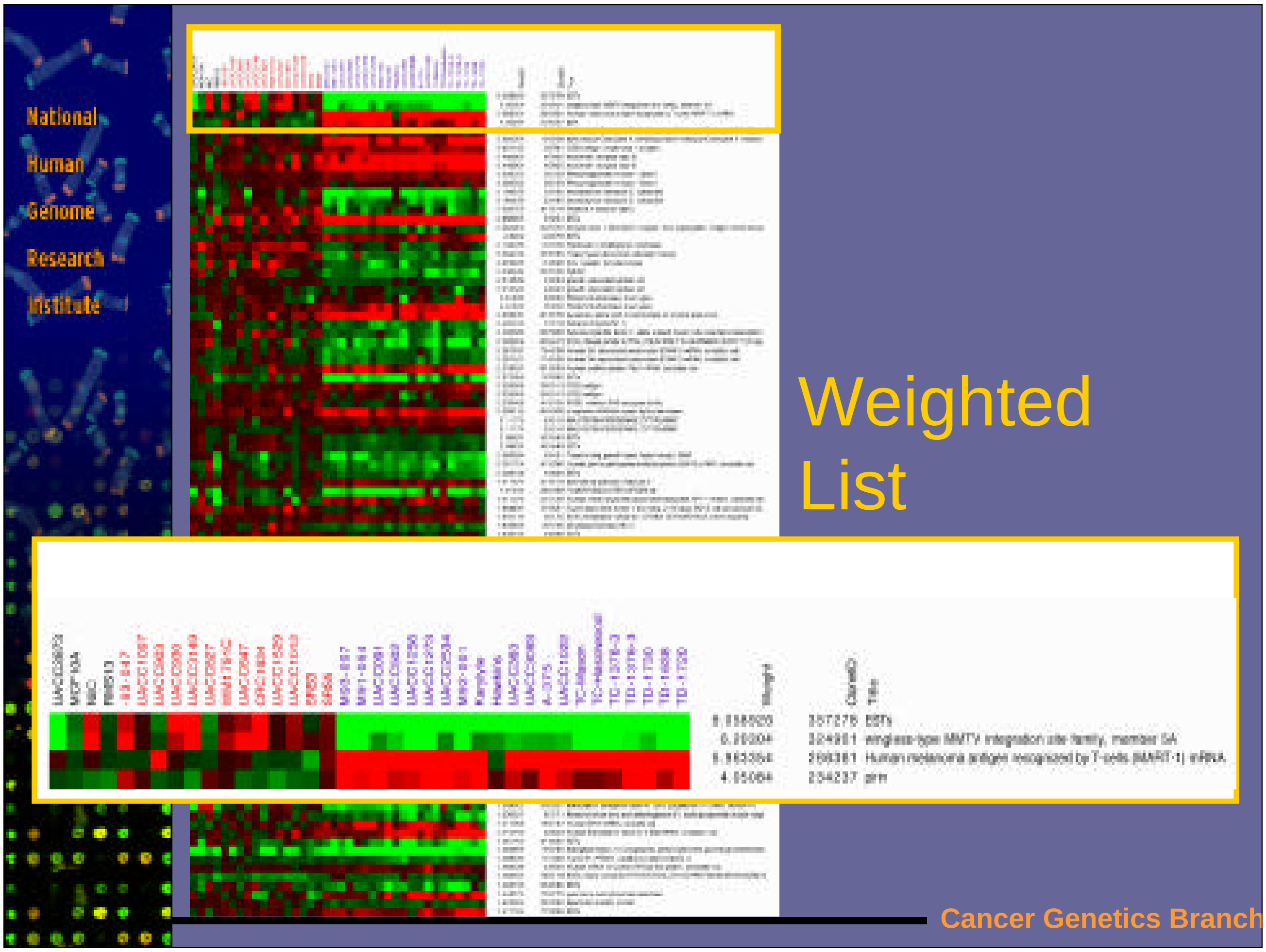
National
Human
Genome
Research
Institute

How subtle a difference in pattern could be reliably detected?

The similarity of the results of the 8 duplicate experiments would put all of these samples within the volume of the central blue sphere.

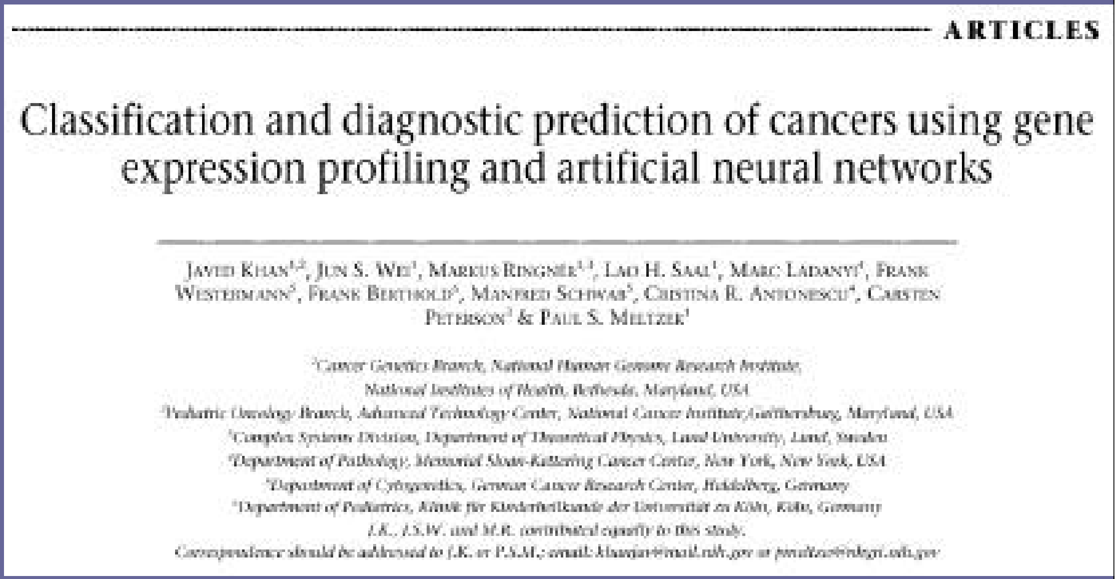
Cancer Genetics Branch

National
Human
Genome
Research
Institute



Weighted List

National
Human
Genome
Research
Institute



ARTICLES

Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks

Javed Khan[1,2], Jun S. Wei[1], Markus Ringnér[1,3], Lao H. Saal[1], Marc Ladanyi[4], Frank Westermann[5], Frank Berthold[6], Manfred Schwab[5], Cristina R. Antonescu[4], Carsten Peterson[3] & Paul S. Meltzer[1]

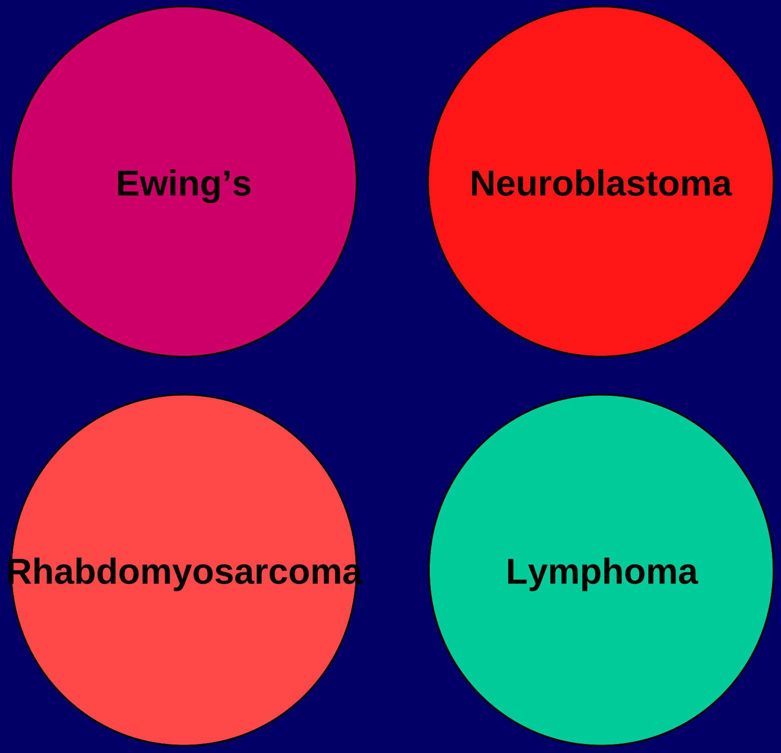
[1]Cancer Genetics Branch, National Human Genome Research Institute,
National Institutes of Health, Bethesda, Maryland, USA
[2]Pediatric Oncology Branch, Advanced Technology Center, National Cancer Institute,Gaithersburg, Maryland, USA
[3]Complex Systems Division, Department of Theoretical Physics, Lund University, Lund, Sweden
[4]Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York, USA
[5]Department of Cytogenetics, German Cancer Research Center, Heidelberg, Germany
[6]Department of Pediatrics, Klinik für Kinderheilkunde der Universität zu Köln, Köln, Germany
J.K., J.S.W. and M.R. contributed equally to this study.
Correspondence should be addressed to J.K. or P.S.M.; email: khanjav@mail.nih.gov or pmeltzer@nhgri.nih.gov

Molecular Taxonomy of Small Round Blue Cell Tumors

Hypothesis

- Using cDNA microarrays we can identify the genes whose expression level is characteristic for that cancer & type
- Utilize these genes to classify the small blue round cell tumors into the correct diagnostic categories

Model: Small Blue Round Cell Tumors



cDNA Microarray Analysis

Experiments

Burkitt's Lymphoma	8
EWS-Tumor	13
EWS-Cell line	10
Neuroblastoma	12
RMS-Tumor	10
RMS-Cell line	10
Test/Unknown	25

>500, 000 Data Points

cDNA microarray

4,000 sequence verified known genes

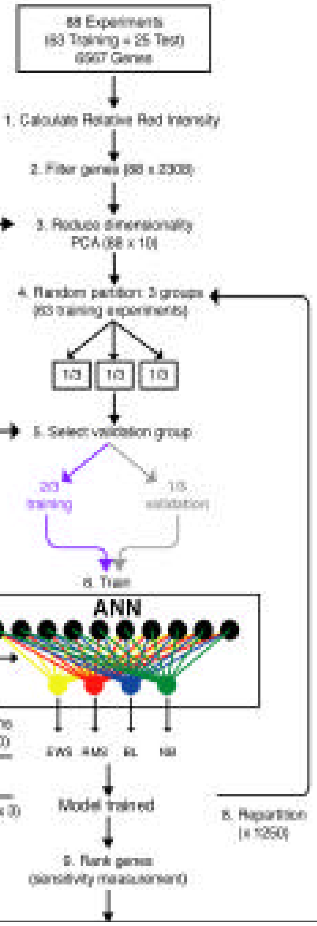
2,567 sequence verified EST

6,567 genes

Artificial Neural Networks

Pattern Recognition

Training



ANN Output
(96 Genes)

Classification 100%

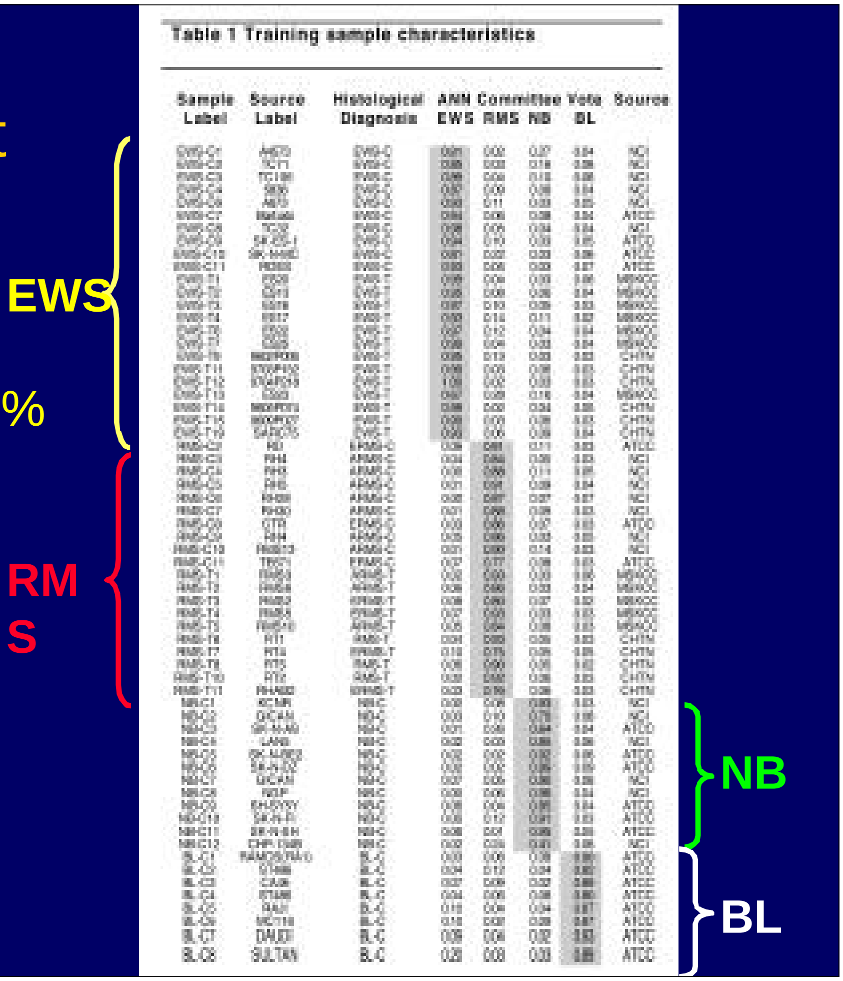
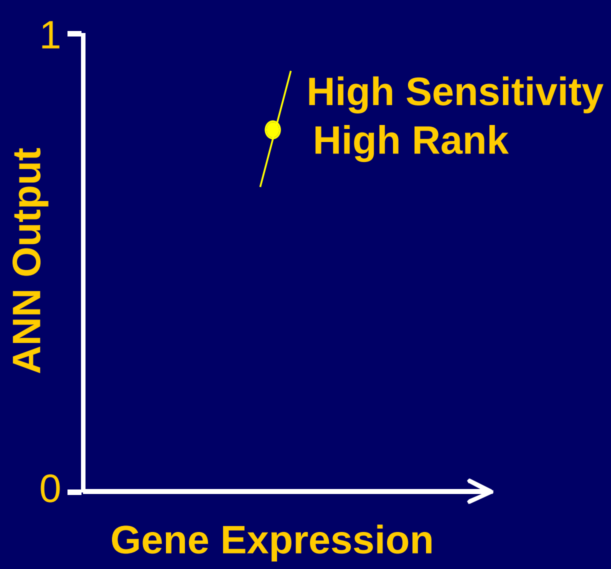

EWS
RMS
NB
BL

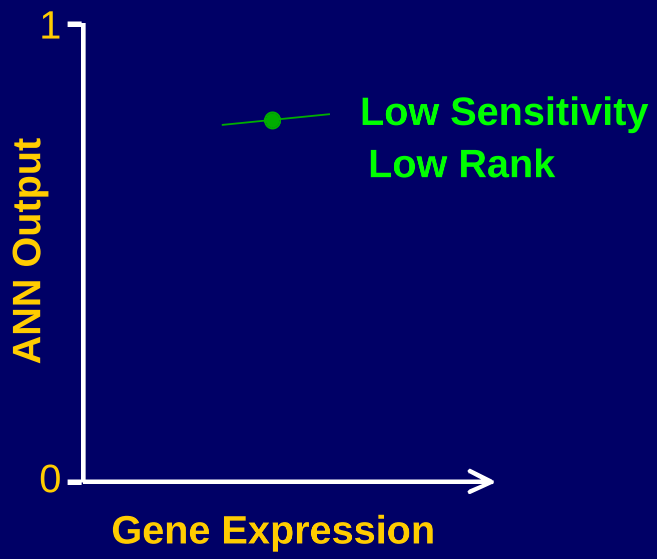
Table 1 Training sample characteristics

Sample Label	Source Label	Histological Diagnosis	ANN Committee Vote EWS	RMS	NB	BL	Source
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Sensitivity Measurement Ranked Genes



Sensitivity Measurement
Ranked Genes


1

0

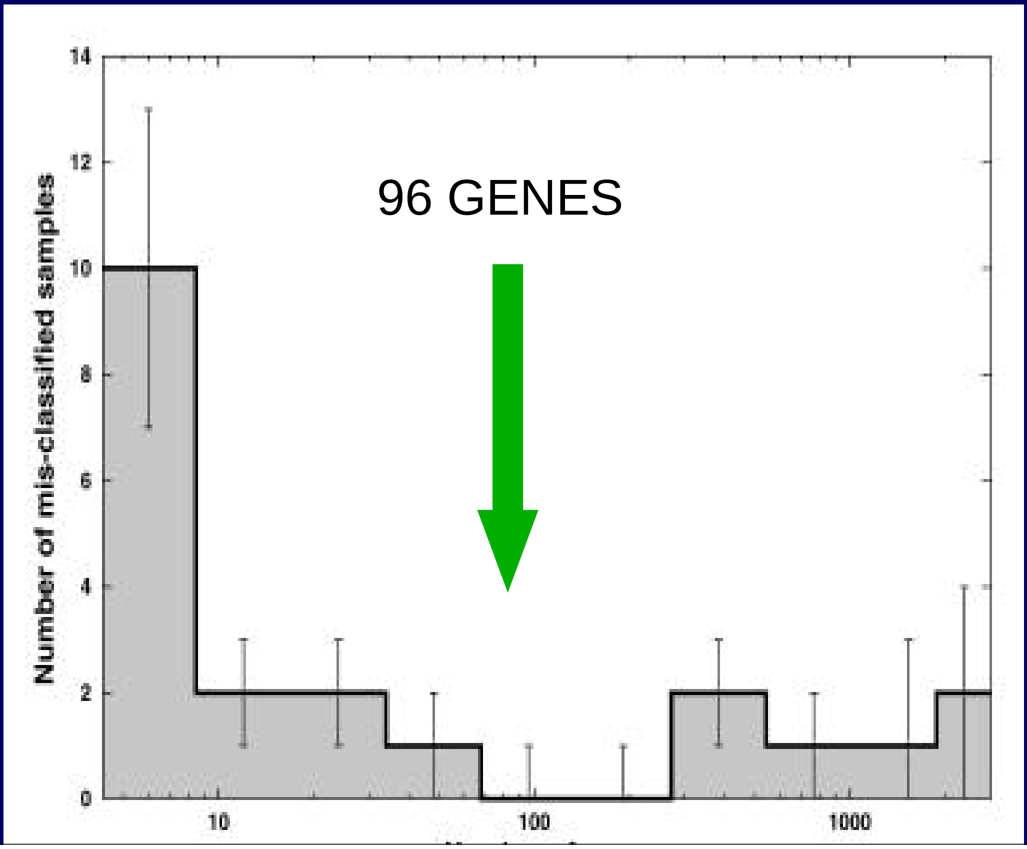
ANN Output

Gene Expression

Low Sensitivity

Low Rank

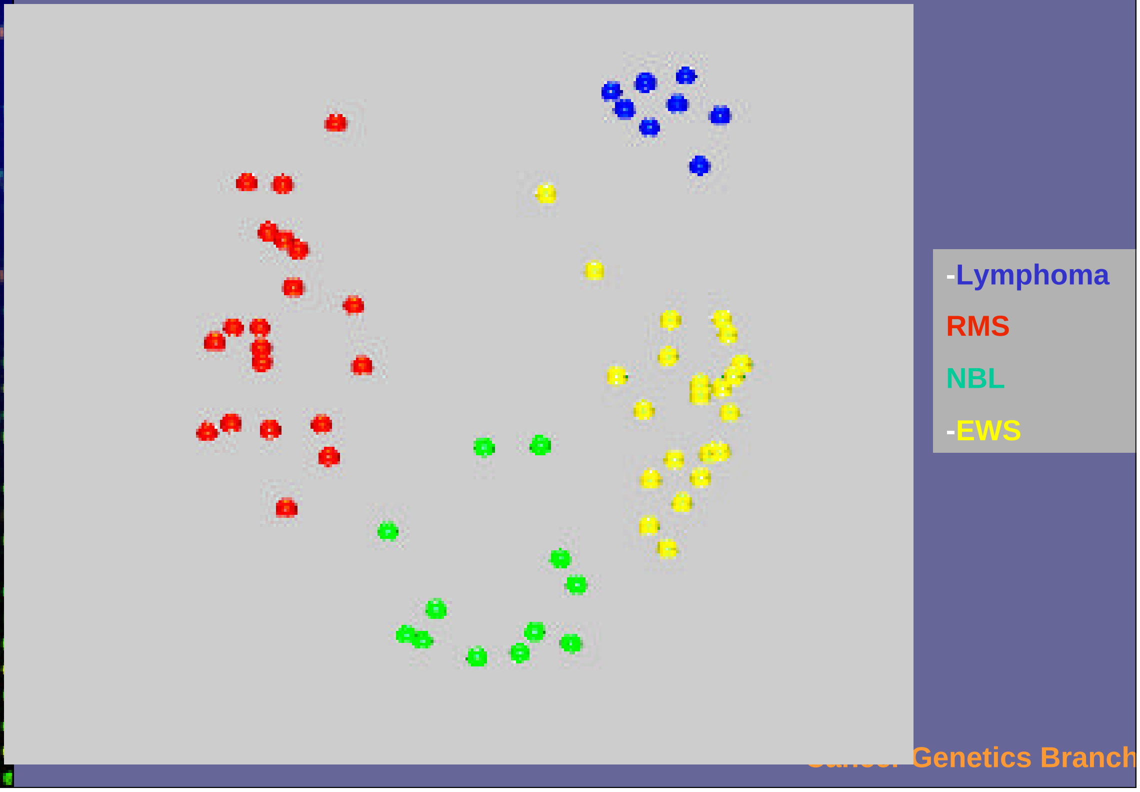
Gene Minimization



96 GENES

National
Human
Genome
Research
Institute

MULTIDIMENSIONAL SCALING

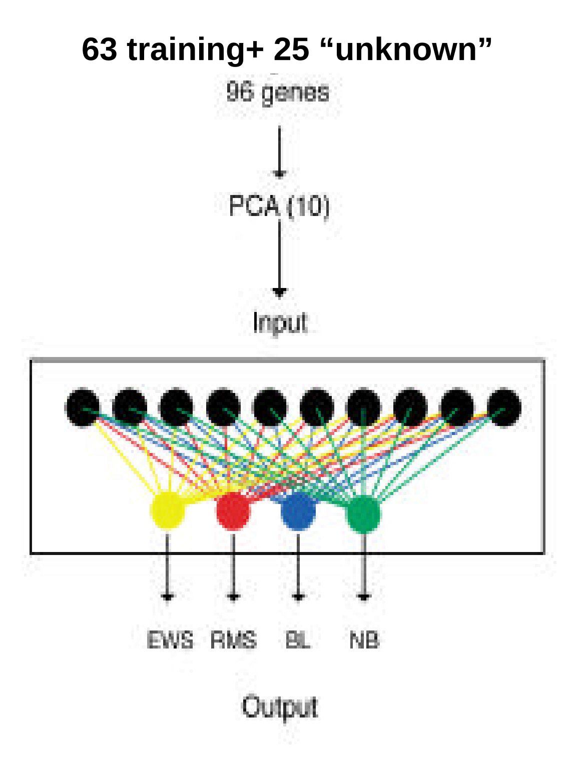


Cancer Genetics Branch

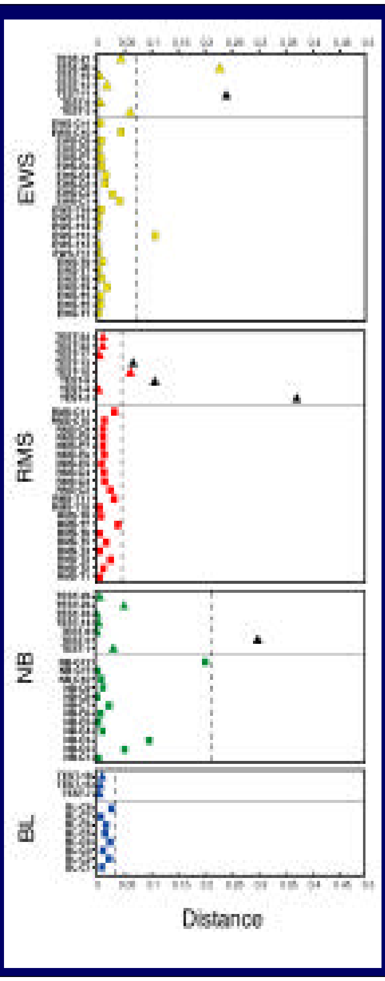
Artificial Neural Networks

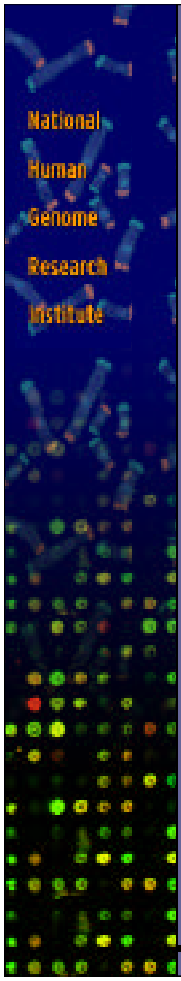
Recalibrate with Top 96 Genes

Diagnosis?


63 training+ 25 “unknown”

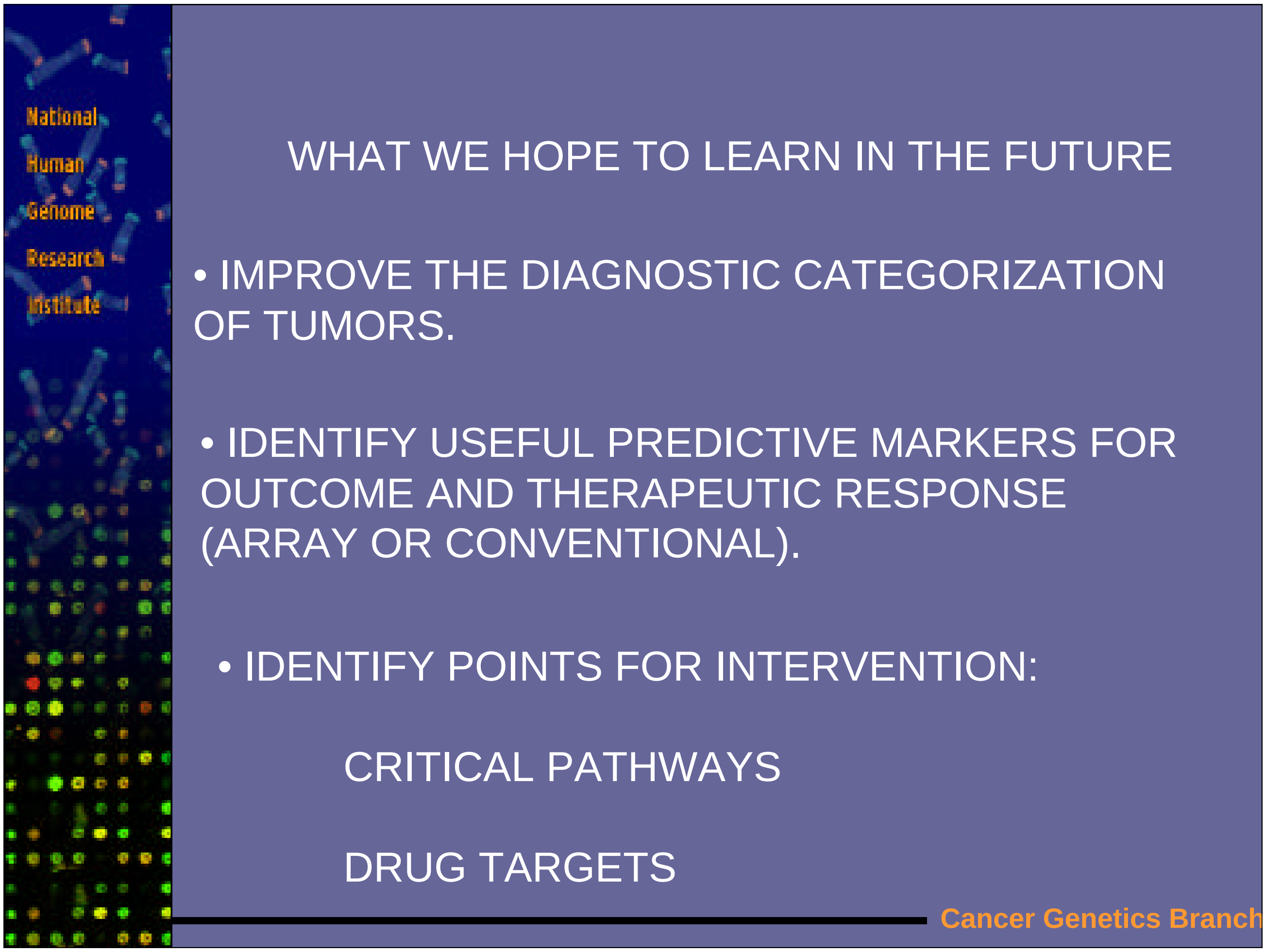
Diagnostic Classification





WHAT HAVE WE LEARNED FROM THE EXPRESSION PROFILING OF CANCERS SO FAR?

- DISTINCT HISTOLOGIES HAVE DISTINCT PATTERNS OF GENE EXPRESSION.
- USING EXPRESSION DATA IT IS POSSIBLE TO DEVELOP ROBUST FORMAL DIAGNOSTIC CLASSIFIERS.
- NOVEL SUBGROUPS CAN BE RECOGNIZED WITHOUT PREVIOUSLY DEFINED HISTOPATHOLOGIC CORRELATES.
- CLINICALLY USEFUL CATEGORIES CAN BE DEFINED, BUT DO NOT NECESSARILY REQUIRE ARRAYS FOR EVERYDAY CLINICAL IDENTIFICATION.



National
Human
Genome
Research
Institute

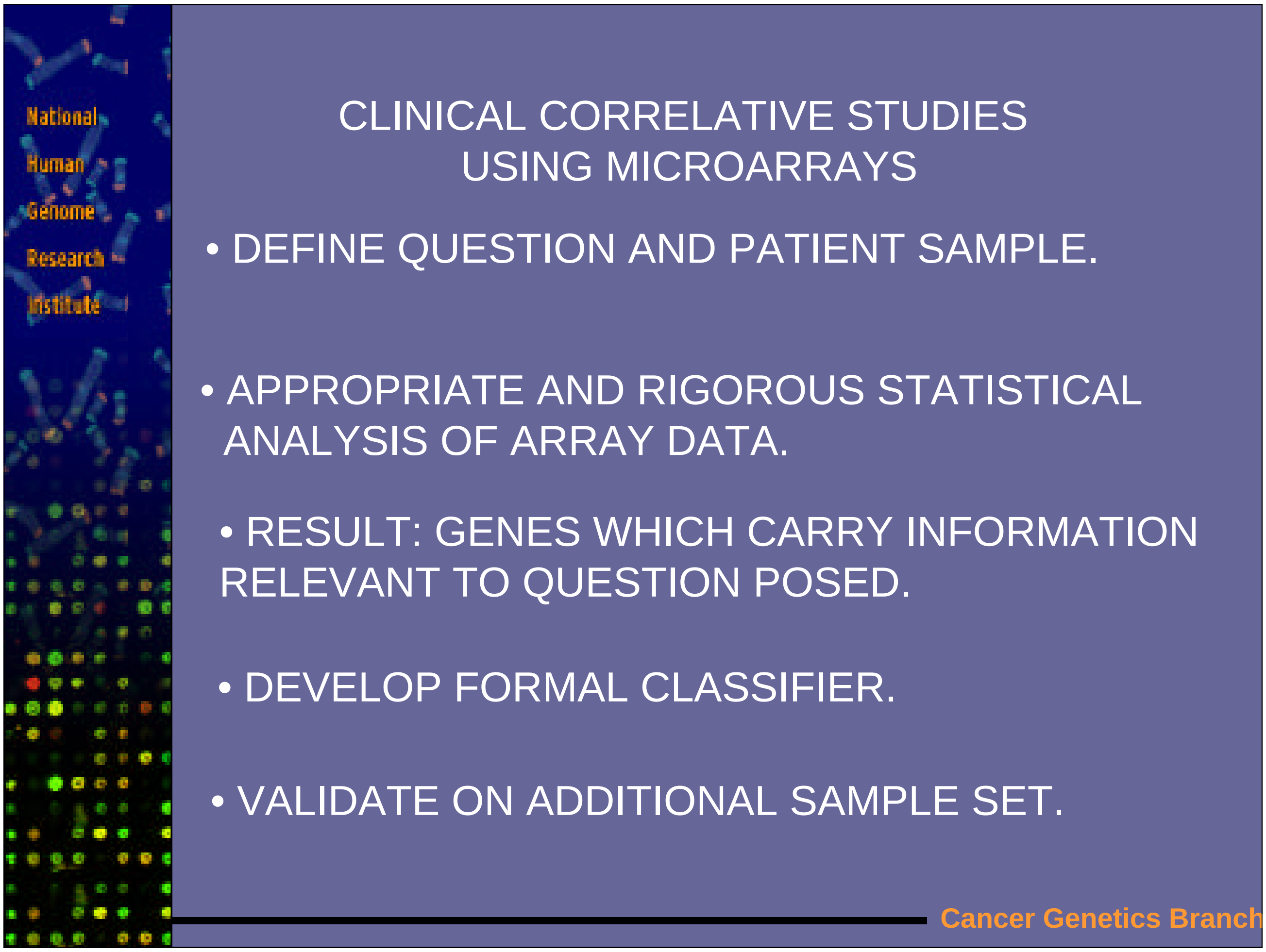
WHAT WE HOPE TO LEARN IN THE FUTURE

- IMPROVE THE DIAGNOSTIC CATEGORIZATION OF TUMORS.
- IDENTIFY USEFUL PREDICTIVE MARKERS FOR OUTCOME AND THERAPEUTIC RESPONSE (ARRAY OR CONVENTIONAL).
- IDENTIFY POINTS FOR INTERVENTION:

  CRITICAL PATHWAYS

  DRUG TARGETS

Cancer Genetics Branch



National
Human
Genome
Research
Institute

CLINICAL CORRELATIVE STUDIES USING MICROARRAYS

- DEFINE QUESTION AND PATIENT SAMPLE.
- APPROPRIATE AND RIGOROUS STATISTICAL ANALYSIS OF ARRAY DATA.
- RESULT: GENES WHICH CARRY INFORMATION RELEVANT TO QUESTION POSED.
- DEVELOP FORMAL CLASSIFIER.
- VALIDATE ON ADDITIONAL SAMPLE SET.

Cancer Genetics Branch

National
Human
Genome
Research
Institute

MODEL SYSTEM WITH CLEAR THERAPEUTIC IMPLICATIONS:
GASTROINTESTINAL STROMAL TUMOR

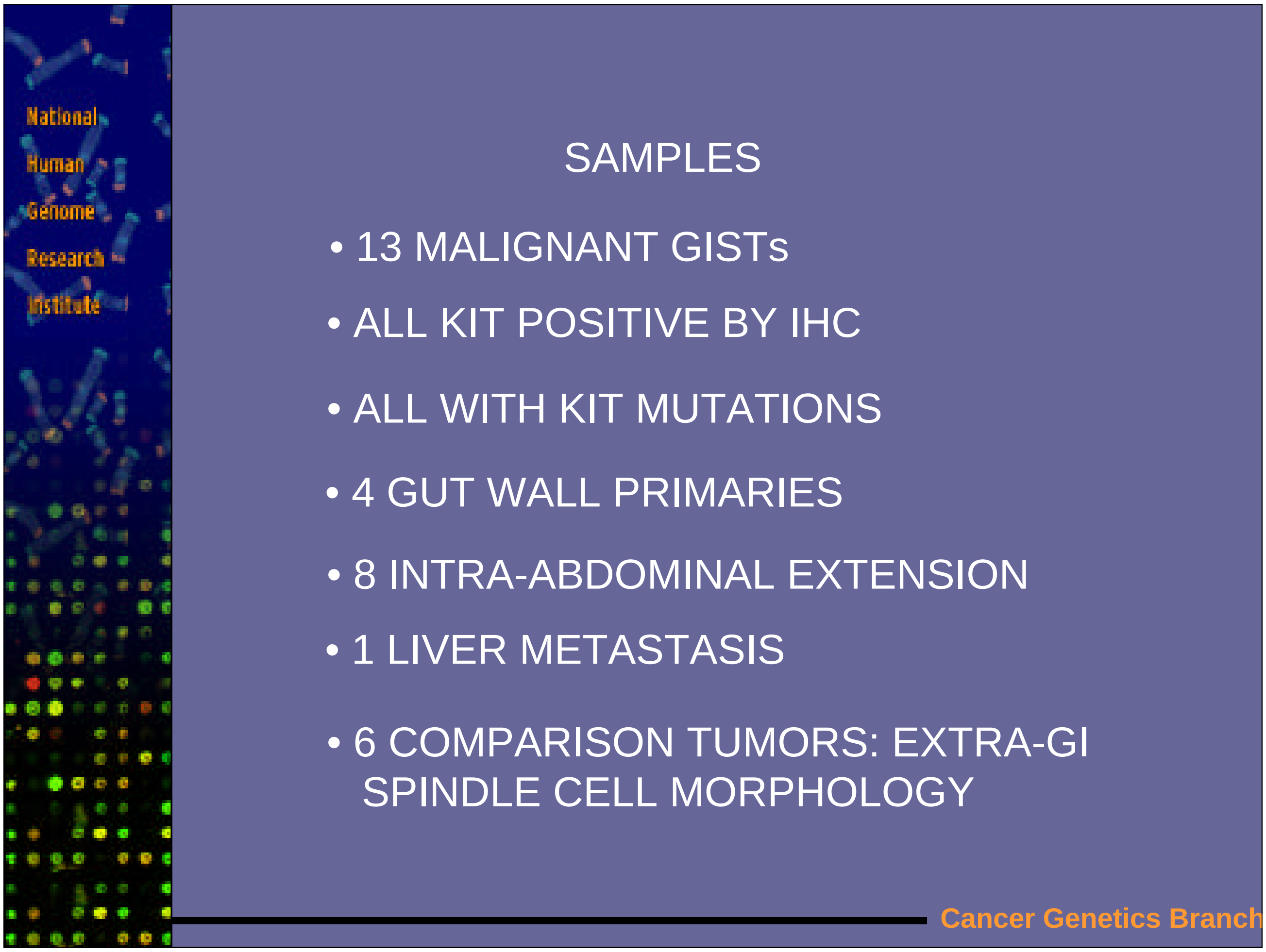
- RELATED TO THE INTERSTITIAL CELLS OF CAJAL

- KIT MUTATIONS

- STI-571 SENSITIVITY

- THE BEST “CREDENTIALED” TARGETS ARE THOSE ACTIVATED BY MUTATION.

Cancer Genetics Branch



National
Human
Genome
Research
Institute

SAMPLES

- 13 MALIGNANT GISTs
- ALL KIT POSITIVE BY IHC
- ALL WITH KIT MUTATIONS
- 4 GUT WALL PRIMARIES
- 8 INTRA-ABDOMINAL EXTENSION
- 1 LIVER METASTASIS
- 6 COMPARISON TUMORS: EXTRA-GI SPINDLE CELL MORPHOLOGY

Cancer Genetics Branch

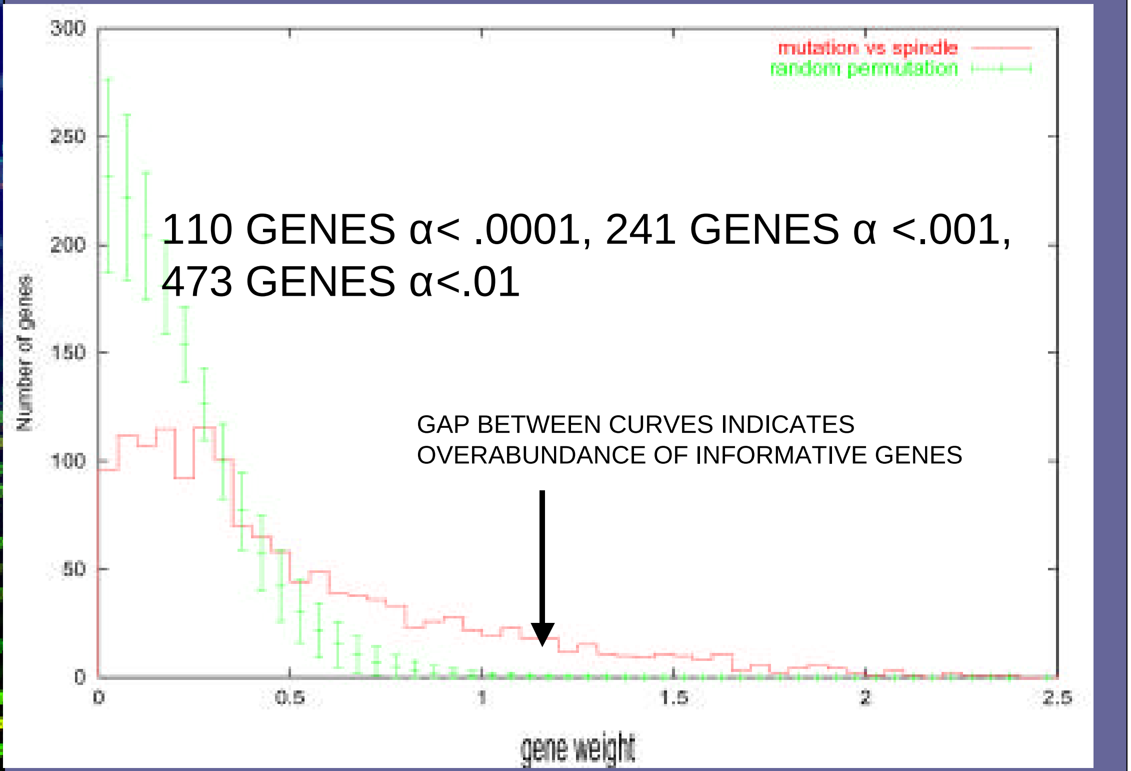
National
Human
Genome
Research
Institute

ARRAYS AND DATA ANALYSIS

- 13,824 ELEMENT SPOTTED cDNA ARRAYS
- OSA REFERENCE PROBE
- GENES RANKED FOR EXPRESSION IN GIST
- WEIGHTED DISCRIMINATOR LIST
- MDS AND HIERARCHICAL CLUSTERING

National
Human
Genome
Research
Institute

OVERABUNDANCE OF INFORMATIVE GENES



110 GENES α< .0001, 241 GENES α <.001, 473 GENES α<.01

GAP BETWEEN CURVES INDICATES OVERABUNDANCE OF INFORMATIVE GENES

National
Human
Genome
Research
Institute

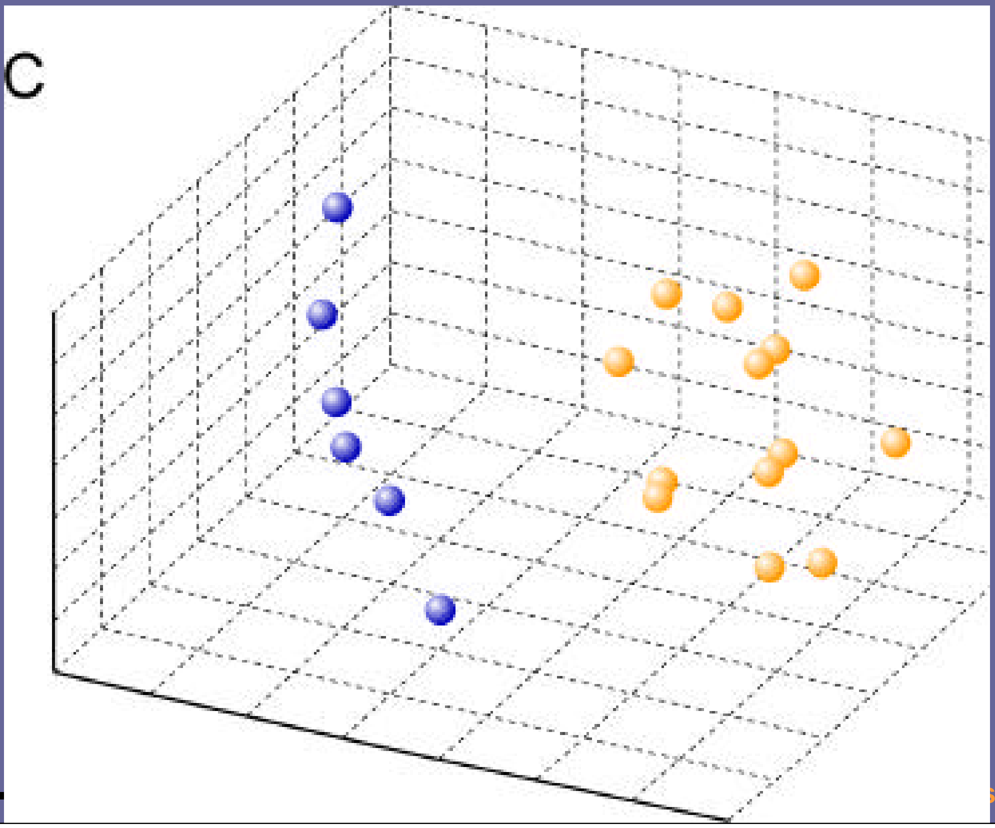
DATA ANALYSIS

- PREFILTER FOR QUALITY AND IDENTIFY GENES HIGHLY EXPRESSED IN GIST: 1987 GENES
- RANK BY WEIGHTED DISCRIMINATOR METHOD
  $w(g, \pm) = ½\mu_+(g)-\mu_-(g)½⁄[\sigma_+(g)+\sigma_-(g)]$
- RANDOM PERMUTATION TEST (10^5 trials)
- CLUSTER ANALYSIS USING DISCRIMINATORS

Cancer Genetics Branch

National
Human
Genome
Research
Institute

MDS PLOT



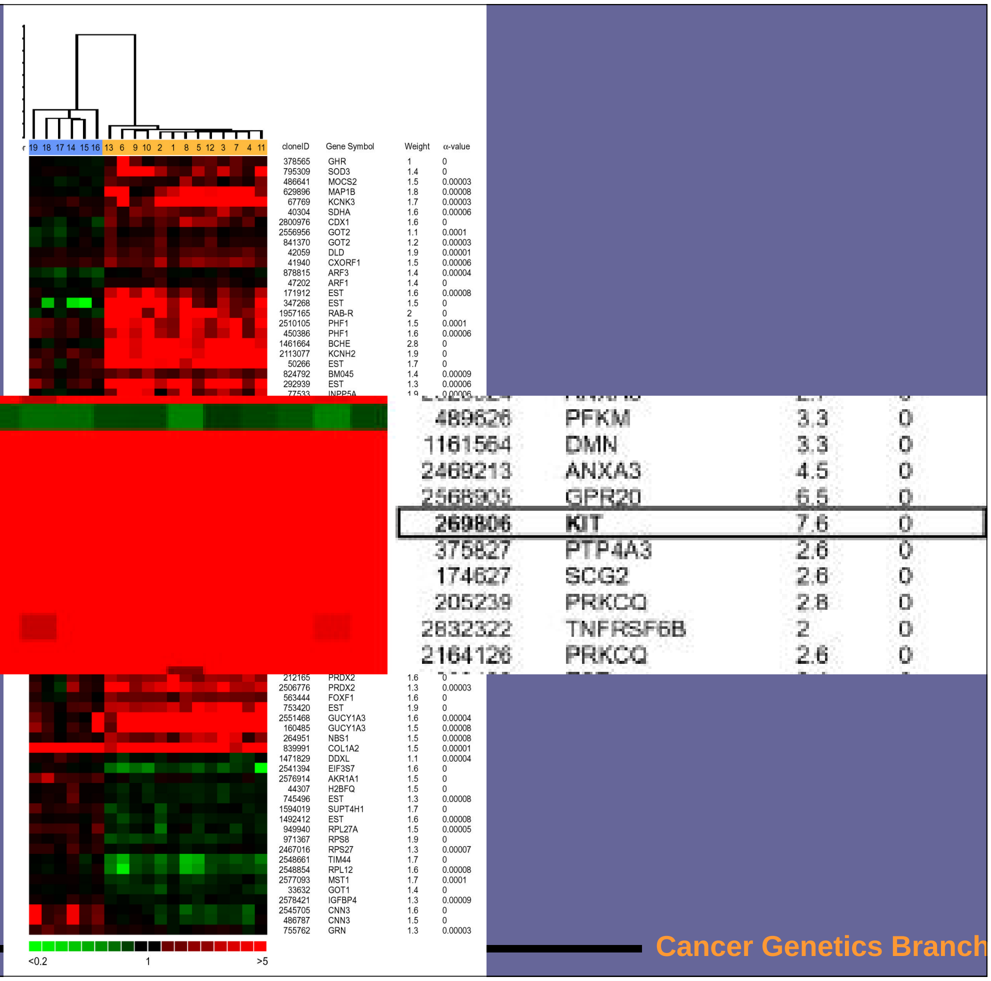
s Branch

National
Human
Genome
Research
Institute

TOP DISCRIMINATORS FOR GIST

Rank	Weight	Alpha	Gene Description
1	7.55575	0	v-kit sarcoma oncogene
2	6.48306	0	G coupled receptor 20
3	4.60057	0	G coupled receptor 20
4	4.51681	0	annexin A3
5	3.33057	0	KIAA0353 protein
6	3.31734	0	phosphofructokinase, muscle
7	2.95095	0.00008	DKFZP434N161 protein
8	2.83435	0	protein kinase C, theta
9	2.79721	0	butyrylcholinesterase
10	2.72752	0	annexin A3

National
Human
Genome
Research
Institute



cloneID	Gene Symbol	Weight	α-value
378565	GHR	1	0
795309	SOD3	1.4	0
486641	MOCS2	1.5	0.00003
629896	MAP1B	1.8	0.00008
67769	KCNK3	1.7	0.00003
40304	SDHA	1.6	0.00006
2800976	CDX1	1.6	0
2556956	GOT2	1.1	0.0001
841370	GOT2	1.2	0.00003
42059	DLD	1.9	0.00001
41940	CXORF1	1.5	0.00006
878815	ARF3	1.4	0.00004
47202	ARF1	1.4	0
171912	EST	1.6	0.00008
347268	EST	1.5	0
1957165	RAB-R	2	0
2510105	PHF1	1.5	0.0001
450386	PHF1	1.6	0.00006
1461664	BCHE	2.8	0
2113077	KCNH2	1.9	0
50266	EST	1.7	0
824792	BM045	1.4	0.00009
292939	EST	1.3	0.00006
77533	INPP5A	1.9	0.00006

489626	PFKM	3.3	0
1161564	DMN	3.3	0
2469213	ANXA3	4.5	0
2568905	GPR20	6.5	0
269806	KIT	7.6	0
375827	PTP4A3	2.6	0
174627	SCG2	2.6	0
205239	PRKCQ	2.8	0
2832322	TNFRSF6B	2	0
2164126	PRKCQ	2.6	0

212165	PRDX2	1.6	0
2506776	PRDX2	1.3	0.00003
563444	FOXF1	1.6	0
753420	EST	1.9	0
2551468	GUCY1A3	1.6	0.00004
160485	GUCY1A3	1.5	0.00008
264951	NBS1	1.5	0.00008
839991	COL1A2	1.5	0.00001
1471829	DDXL	1.1	0.00004
2541394	EIF3S7	1.6	0
2576914	AKR1A1	1.5	0
44307	H2BFQ	1.5	0
745496	EST	1.3	0.00008
1594019	SUPT4H1	1.7	0
1492412	EST	1.6	0.00008
949940	RPL27A	1.5	0.00005
971367	RPS8	1.9	0
2467016	RPS27	1.3	0.00007
2548661	TIM44	1.7	0
2548854	RPL12	1.6	0.00008
2577093	MST1	1.7	0.0001
33632	GOT1	1.4	0
2578421	IGFBP4	1.3	0.00009
2545705	CNN3	1.6	0
486787	CNN3	1.5	0
755762	GRN	1.3	0.00003

CONCLUSIONS

- MALIGNANT GISTS EXHIBIT A DISTINCT AND HIGHLY COHERENT GENE EXPRESSION PROFILE.

- THIS GENE LIST IS RELEVANT BOTH TO GIST GROWTH AND NORMAL ICC FUNCTION.

- KIT, A CRITICAL GENE IN REGULATING GIST GROWTH, IS THE BEST DISCRIMINATOR FOR THIS DISEASE.

- EXTENDING THIS APPROACH TO OTHER CANCERS MAY HELP IDENTIFY NEW DISEASE SPECIFIC DRUG TARGETS.

A RECURRING PROBLEM

Oncogenes

Transcription factors

Hormones/growth factors

Drugs

Toxins

Radiation

?????

<u>Downstream Genes</u>

•Direct targets

•Indirect targets

Retrovirus:

Empty vector

Pax3

Pax3-Fkhr

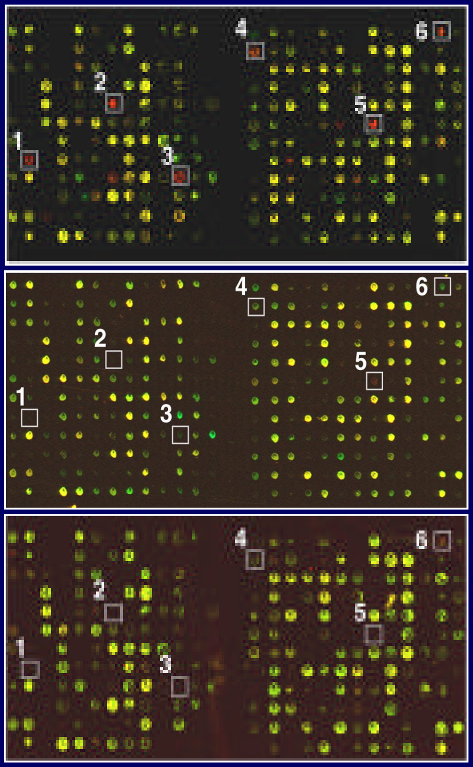
RESULTS

Mouse cDNA array
2200 genes

PAX3-FKHR
vs
Empty Vector

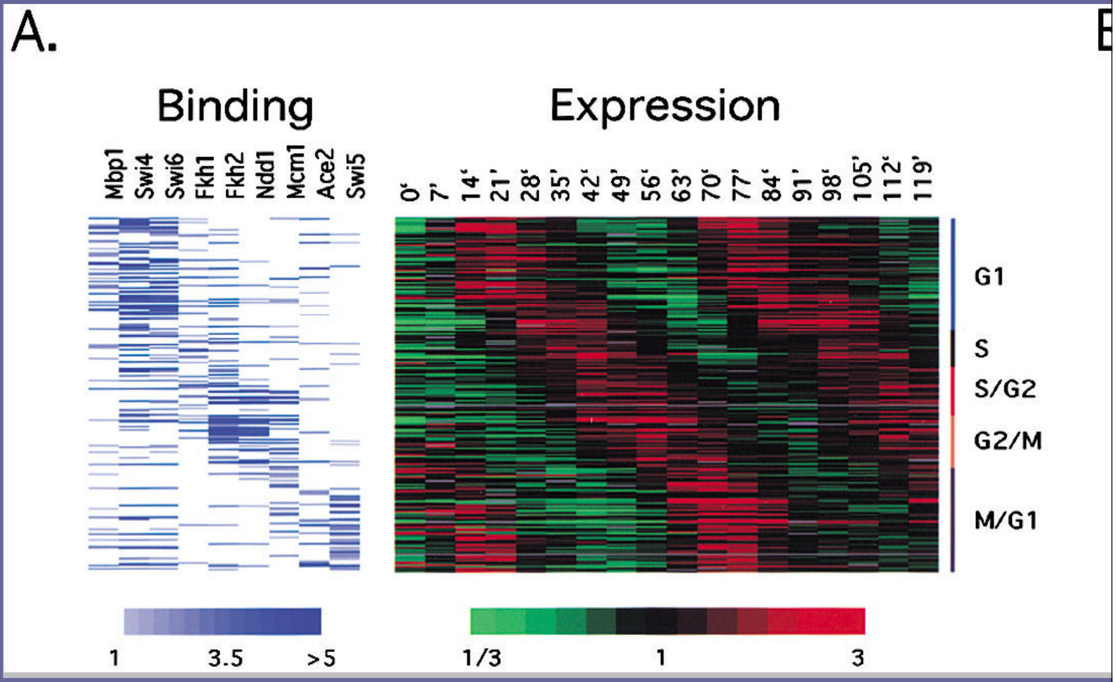
PAX3
vs
Empty Vector

3T3 PARENT
vs
Empty Vector



1. Troponin C
2. IGFBP5
3. Myogenin
4. Six1
5. Troponin T
6. IGF2

Promoter Occupancy During Yeast Cell Cycle



Simon I Cell. 2001 Sep 21;106(6):697-708.

Future Promise

• Numerous clever applications to additional systems, new
computational tools and technical innovations.

• Development of large clinically annotated datasets which allow precise definition of patient subsets and lead to the identification of new therapeutic targets.

• Linking genomic sequence to expression data to define regulatory elements/transcription factors associated with co-regulated genes.

•Development of computational tools which allow predictive
modeling of gene networks.

•Introduction of technologies which increases the dimensionality of expression data: protein arrays; cell arrays; promoter arrays.